Recent Trends in Skill for Some Leading Global NWP Centers

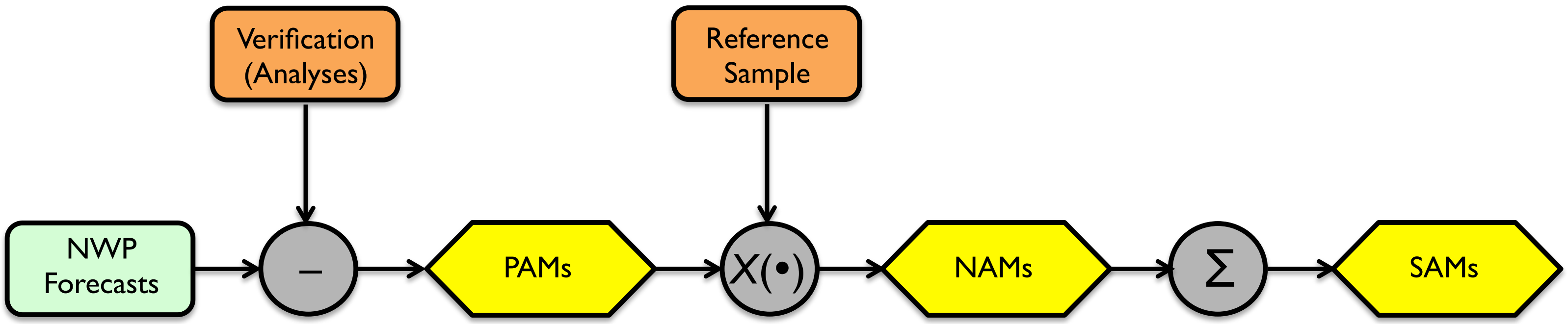
Ross N Hoffman[a,b,d,e], Krishna Kumar[c,d], Sid A. Boukabara[d], Kayo Ide[e],
Fanglin Yang[f], Robert Atlas[a]

7 May 2019

(a) NOAA/Atlantic Oceanographic and Meteorological Laboratory, Miami, Florida
(b) Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, Florida
(c) Riverside Technology Inc., College Park, Maryland
(d) NOAA/NESDIS/STAR, College Park, Maryland
(e) University of Maryland, College Park, College Park, Maryland
(f) NOAA/NCEP/Environmental Modeling Center, College Park, Maryland

Introduction

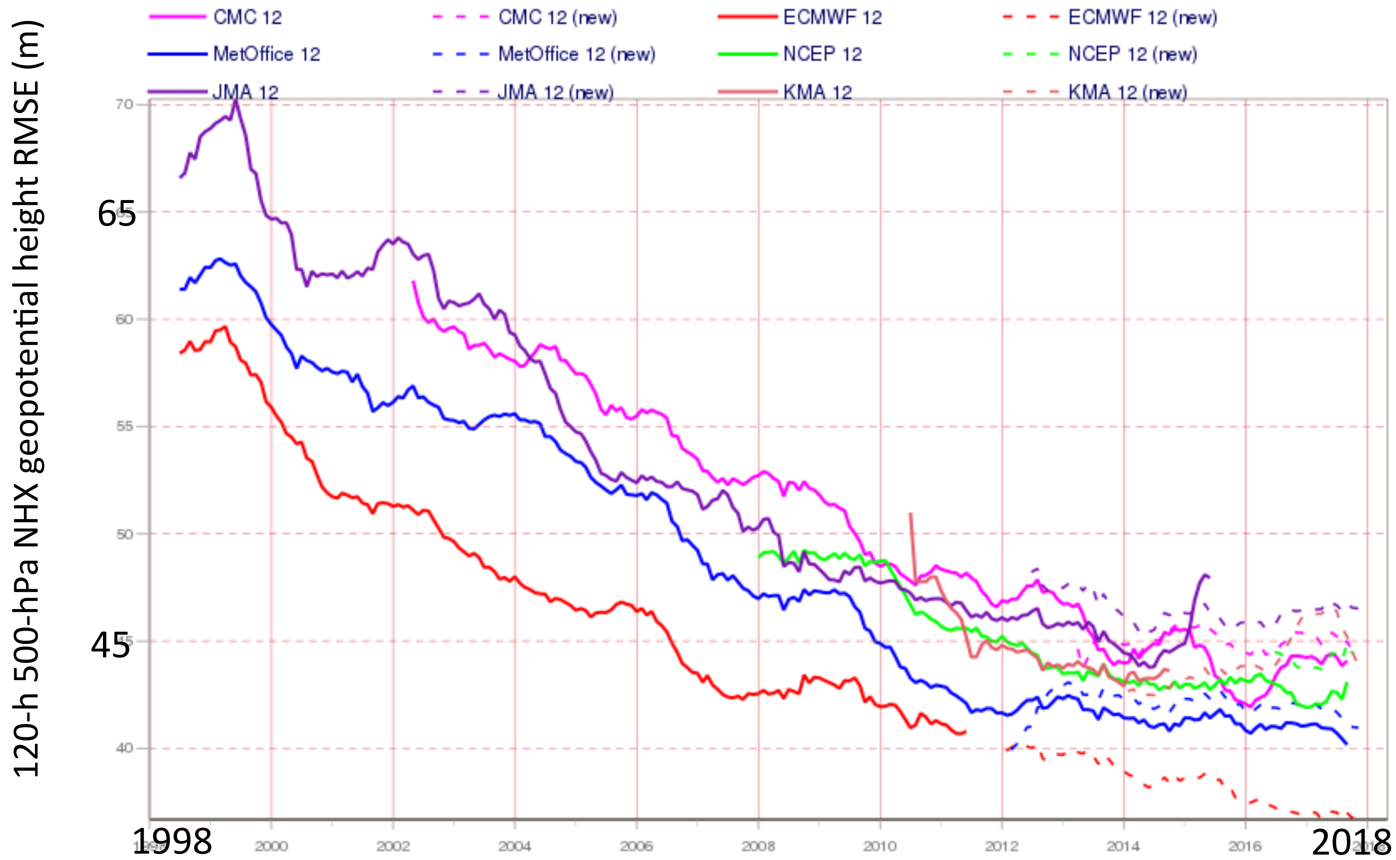
- A look at the deterministic forecasts of three leading NWP centers (ECMWF, NCEP, UKMO) for the years 2015-2017.
- PAMs (primary assessment metrics) such as the 500-hPa geopotential anomaly correlation (AC) or the 250-hPa wind RMSE are converted to NAMs (normalized assessment metrics) and then averaged into SAMs (summary assessment metrics).



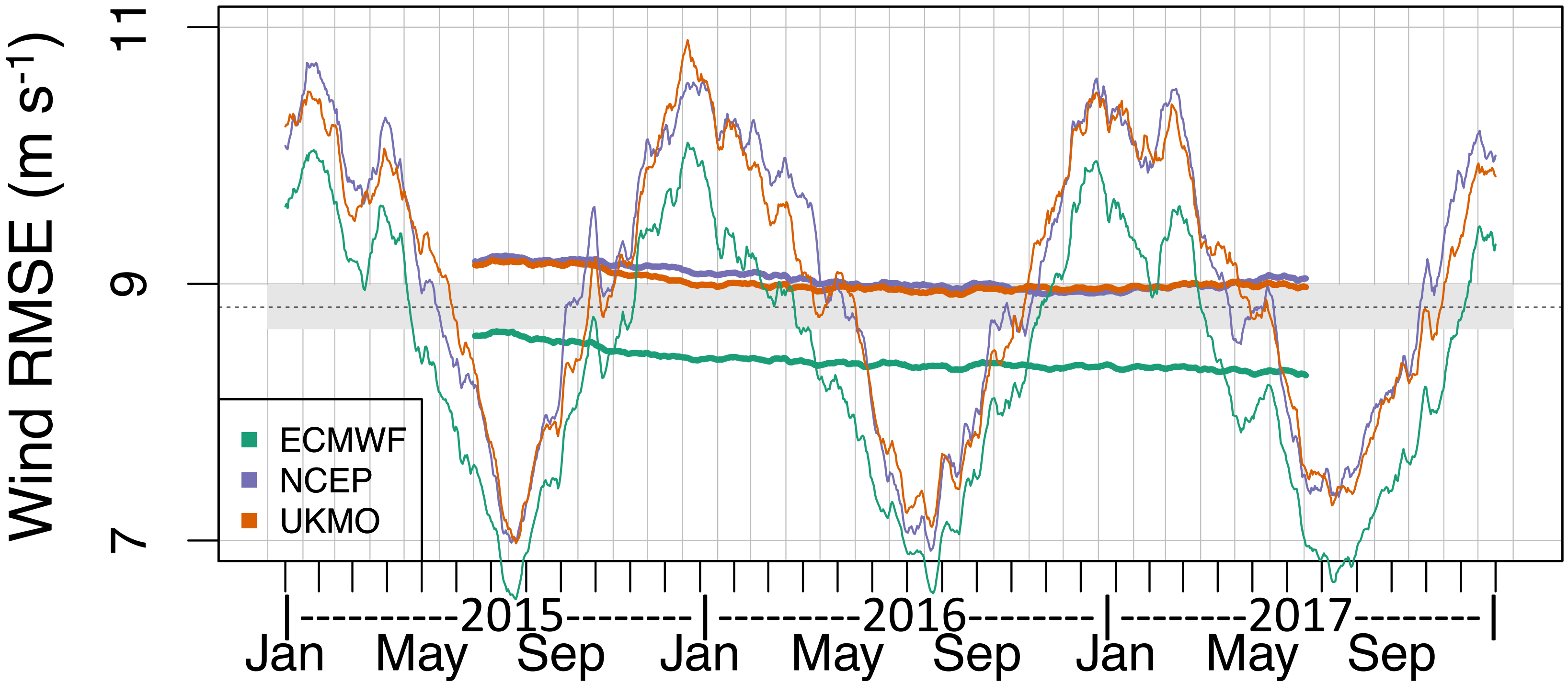
Upgrades 2015-2017

Center	i	Date	Upgrade	Delta
ECMWF	1	20150512	IFS Cycle 41r1	2.10
	2	20160308	IFS Cycle 41r2 (Cubic Octahedral 1280)	1.31
	3	20161122	IFS Cycle 43r1	2.58
	4	20170717	IFS Cycle 43r3	5.22
NCEP	1	20150114	TIN14-46 (T1534)	-4.12
	2	20160511	TIN16-11 (4DEnVar)	7.37
	3	20170719	SCN17-67 (NEMSIO)	0.81
UKMO	1	20161121	PS38 (satellite obs.)	4.75
	2	20170907	PS39 (10-km resolution)	2.82

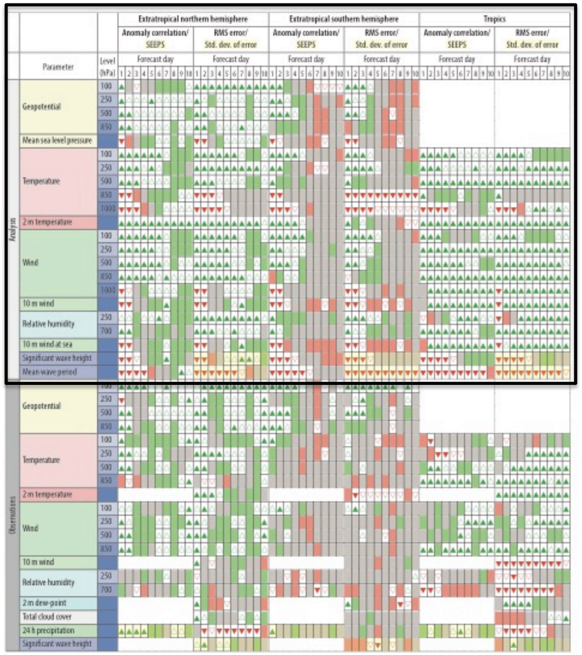
Context: 20 years of forecast skill



120-h 500-hPa NHX vector wind RMSE; MA(365) and MA(31)



Scorecards of IFS Cycle 45r1 versus IFS Cycle 43r3.
From ECMWF Newletter No. 156.



Symbol legend: for a given forecast step...

- ▲ 45r1 better than 43r3 statistically significant with 99.7% confidence
- △ 45r1 better than 43r3 statistically significant with 95% confidence
- 45r1 better than 43r3 statistically significant with 68% confidence
- no significant difference between 43r3 and 45r1
- 45r1 worse than 43r3 statistically significant with 68% confidence
- ▽ 45r1 worse than 43r3 statistically significant with 95% confidence
- ▼ 45r1 worse than 43r3 statistically significant with 99.7% confidence

Figure 1 HRES scorecard of IFS Cycle 45r1 versus IFS Cycle 43r3, verified by the respective analyses and observations at 00 and 12 UTC, based on 855 forecast runs in the period December 2016 to June 2018. See Box A for a discussion of how scores computed against analyses have been affected by changes to the analysis in IFS Cycle 45r1.

reduced spread in clear-sky regions (due to unperturbed radiative tendency in clear sky), the activation of tendency perturbations in the stratosphere, and weaker tapering of perturbations in the boundary layer; a reduction in the amplitude of the SPPT perturbation patterns (by 20%); introduction of the cycling of stochastic physics random fields in the EDA, and adoption of the same SPPT configuration in EDA as in ENS; deactivation of the stochastic backscatter (SKEB) scheme due to improved model error representation by the SPPT scheme (see above), leading to a 2.5% cost saving in ENS.

- **Software infrastructure:** the *ecBuild* system is incorporated into the IFS source repository, which enables a standalone build of the IFS to be created on a workstation, with all required dependencies resolved automatically, and a small quality assurance test suite to be run. This will help to develop and test future code changes more efficiently.

Impacts

A comparison of parallel runs of the previous operational cycle (43r3) and the new cycle (45r1) indicates an overall positive impact in the tropics for both HRES and ENS (Figures 1 and 2). For the extratropics, results are mixed, with an overall slightly positive impact on the HRES scores, while for the ENS the sign of the impact depends on the geographical region and the variable.

Upper-air fields

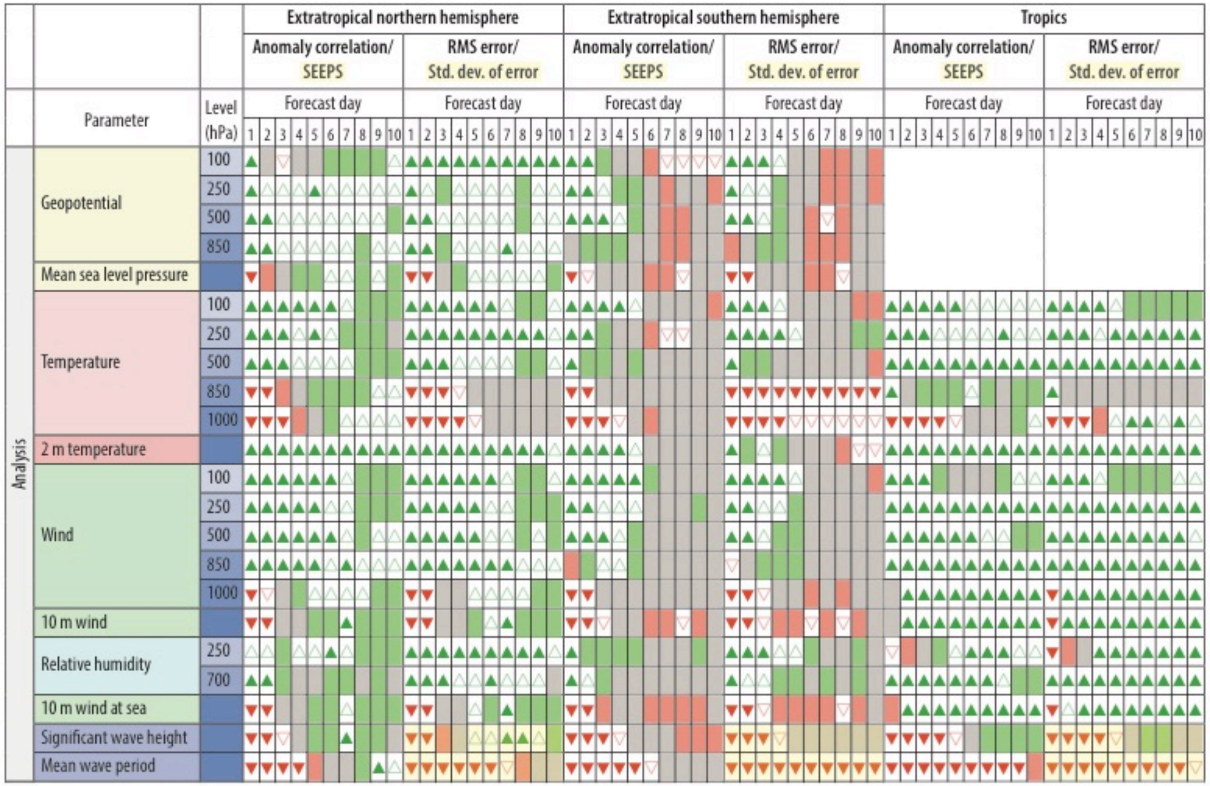
The new cycle leads to improvements in HRES upper-air fields. When these fields are verified against the

Symbol legend: for a given forecast step...

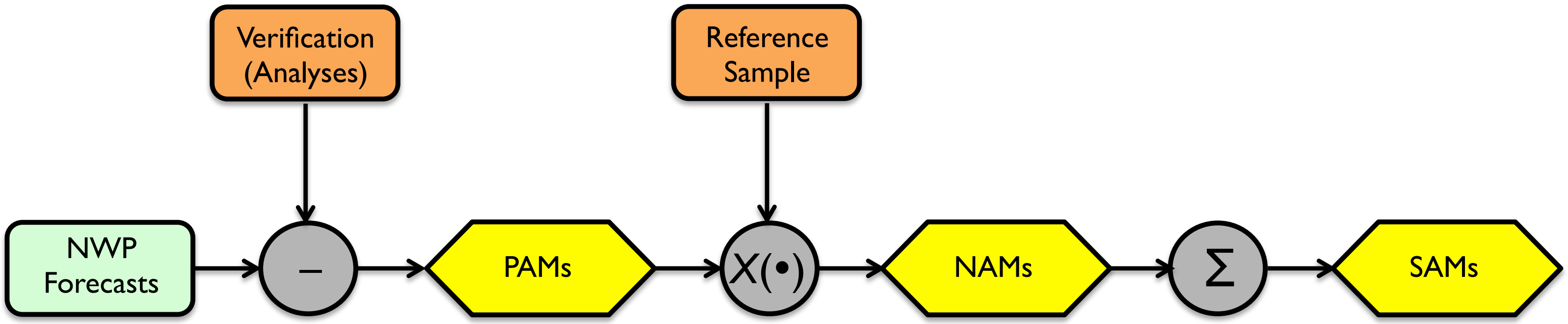
- ▲ 45r1 better than 43r3 statistically significant with 99.7% confidence
- △ 45r1 better than 43r3 statistically significant with 95% confidence
- 45r1 better than 43r3 statistically significant with 68% confidence
- no significant difference between 43r3 and 45r1
- 45r1 worse than 43r3 statistically significant with 68% confidence
- ▽ 45r1 worse than 43r3 statistically significant with 95% confidence
- ▼ 45r1 worse than 43r3 statistically significant with 99.7% confidence

Figure 2 ENS scorecard of IFS Cycle 45r1 versus IFS Cycle 43r3 for medium-range/monthly forecasts up to forecast day 15, verified by the respective analyses and observations at 00 and 12 UTC, based on 408 ENS forecast runs in the period December 2016 to June 2018. See Box A for a discussion of how scores computed against analyses have been affected by changes to the analysis in IFS Cycle 45r1.

Scorecards of IFS Cycle 45r1 versus IFS Cycle 43r3.
From ECMWF Newletter No. 156. Showing HRES vs. analysis only.

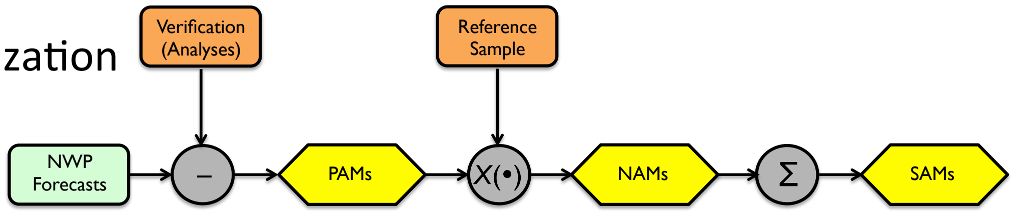


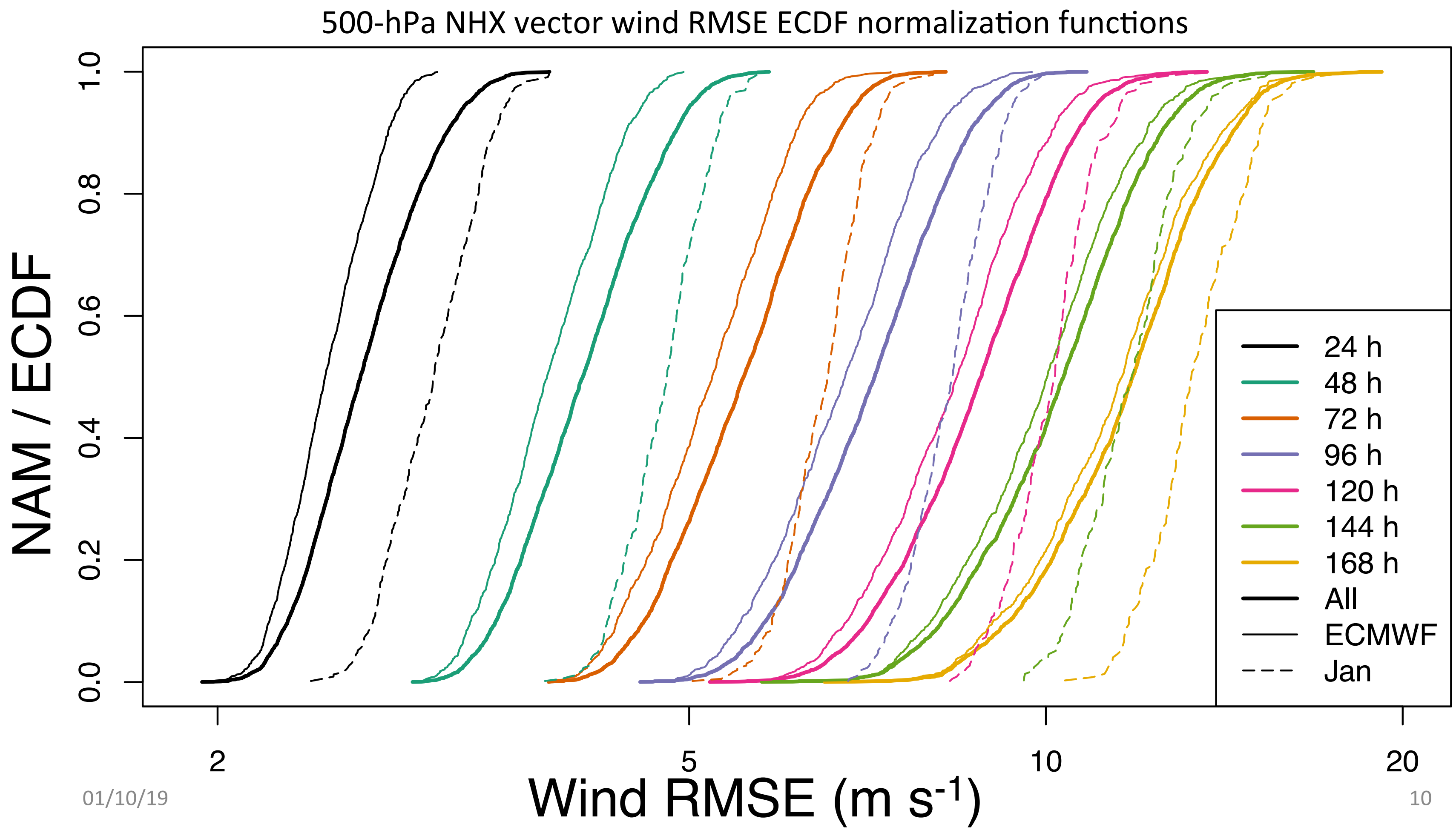
PAMs to NAMs to SAMs

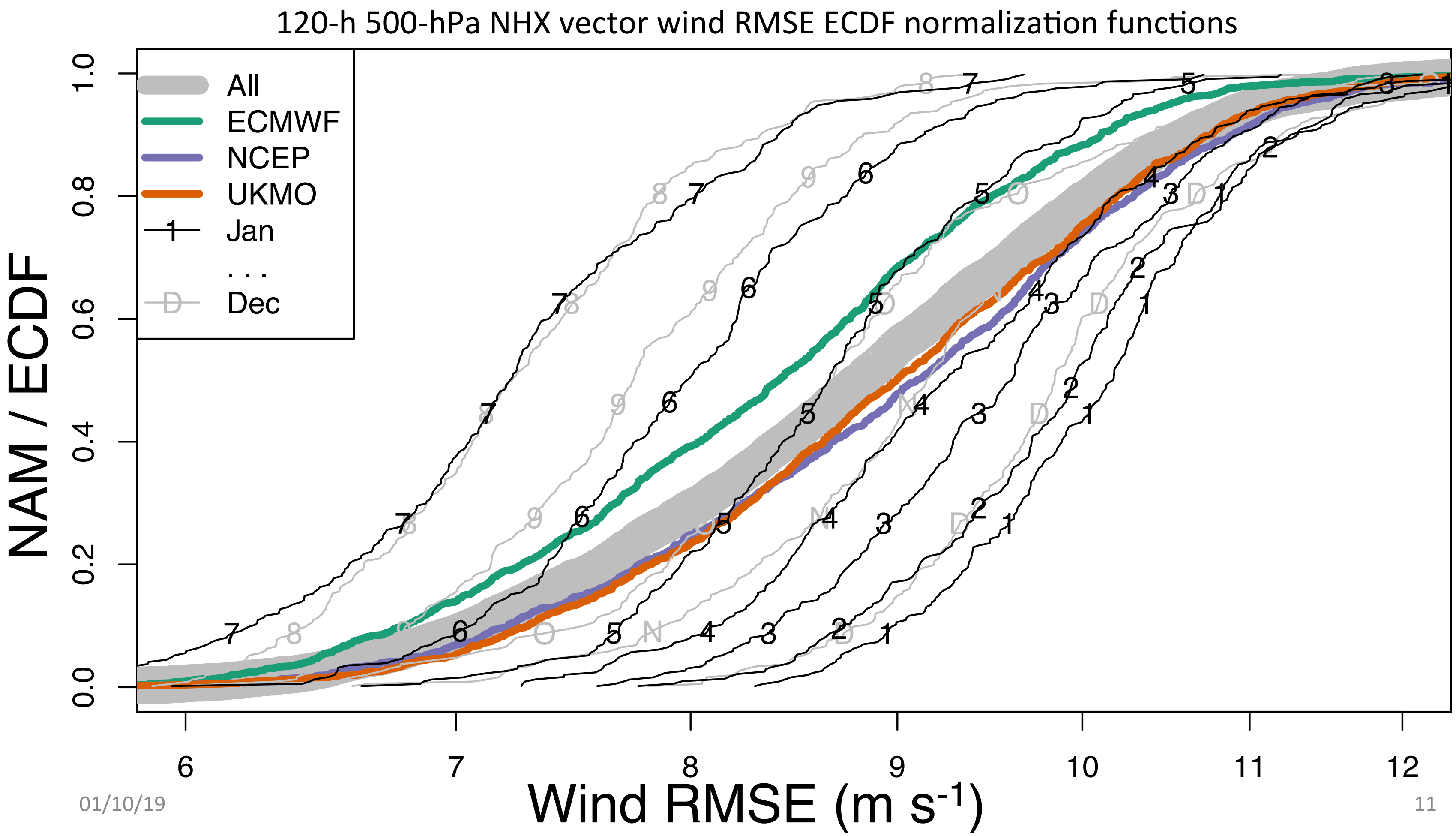


PAMs to NAMs to SAMs

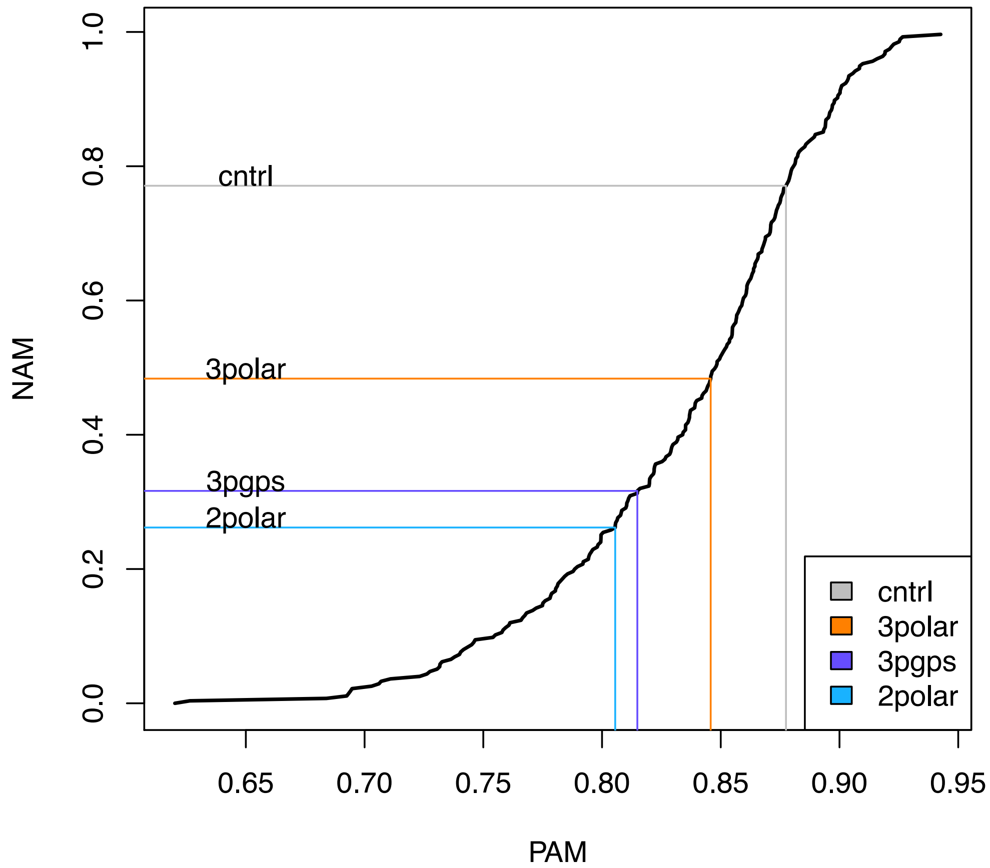
- We often focus on a few key PAMs, but this may ignore other important aspects of forecast skill. The use of SAMs increases statistical significance and enables exploring different aspects of forecast skill.
- PAM/NAM/SAM dimension :: coordinate values
  - Forecast time :: 24, 48, 72, 96, 120, 144, 168 h
  - Level :: 250, 500, 700, 850, 1000 hPa
  - Domain :: northern hemisphere extratropics (NHX), southern hemisphere extratropics (SHX), tropics
  - Variable :: height (Z), temperature (T), wind (V)
  - Statistic :: anomaly correlation (AC), root mean square error (RMSE), absolute mean error (AME, the absolute value of bias)
  - Verification time :: every 24 h at 0000 UTC during 2015-2017
  - Center :: ECMWF, NCEP, UKMO
- Reference sample for normalization
  - All :: (verification time, center)
  - ByCenter :: (verification time)




500-hPa NHX vector wind RMSE ECDF normalization functions
NAM / ECDF
0.0
0.2
0.4
0.6
0.8
1.0
2
5
10
20
Wind RMSE (m s-1)
24 h
48 h
72 h
96 h
120 h
144 h
168 h
All
ECMWF
Jan

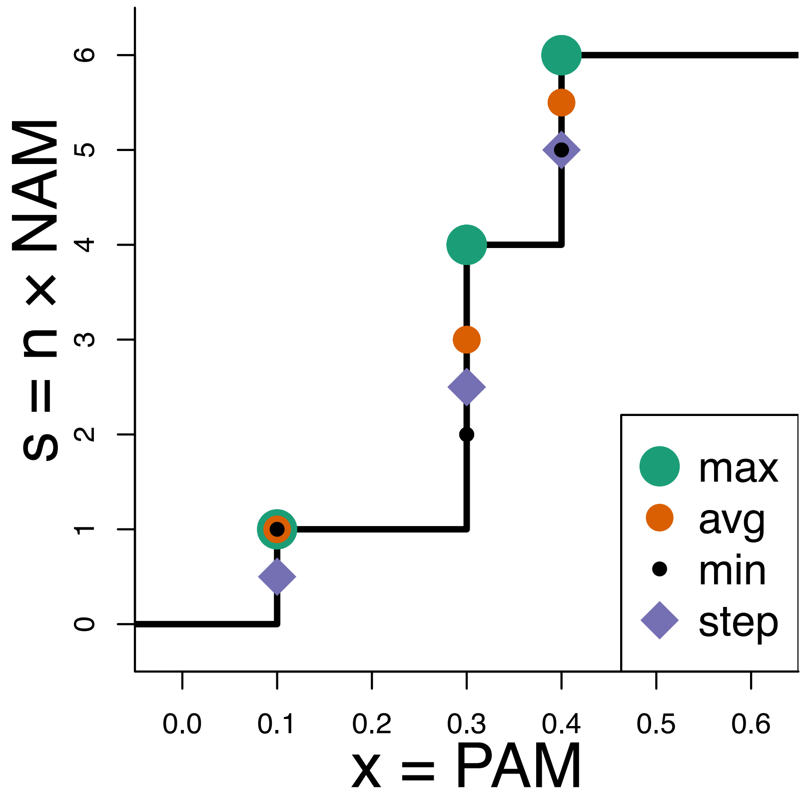

120-h 500-hPa NHX vector wind RMSE ECDF normalization functions
All
ECMWF
NCEP
UKMO
Jan
. . .
Dec
NAM / ECDF
Wind RMSE (m s-1)
0.0
0.2
0.4
0.6
0.8
1.0
6
7
8
9
10
11
12

Example



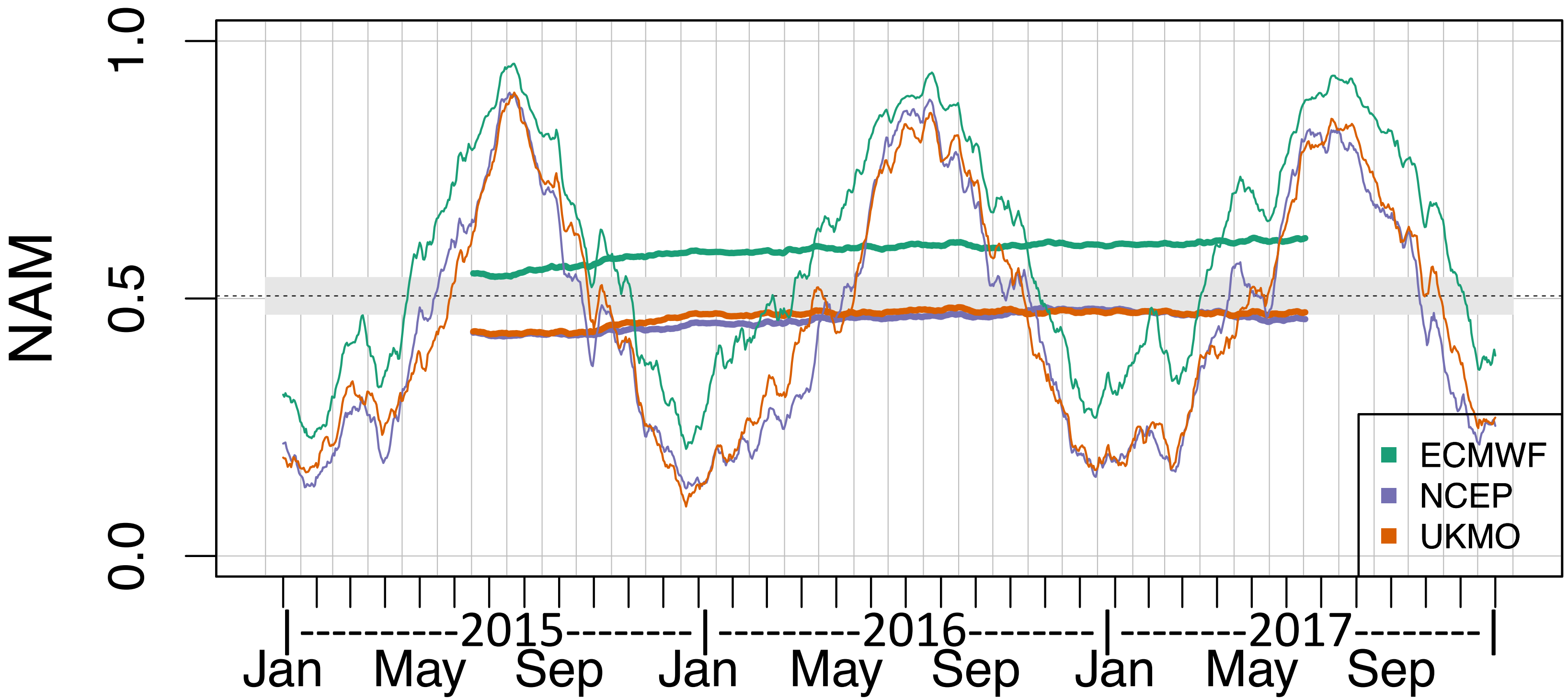
cntrl
3polar
3pgps
2polar
NAM
PAM
0.0
0.2
0.4
0.6
0.8
1.0
0.65
0.70
0.75
0.80
0.85
0.90
0.95
cntrl
3polar
3pgps
2polar

ECDF NAM is calculated from the rank

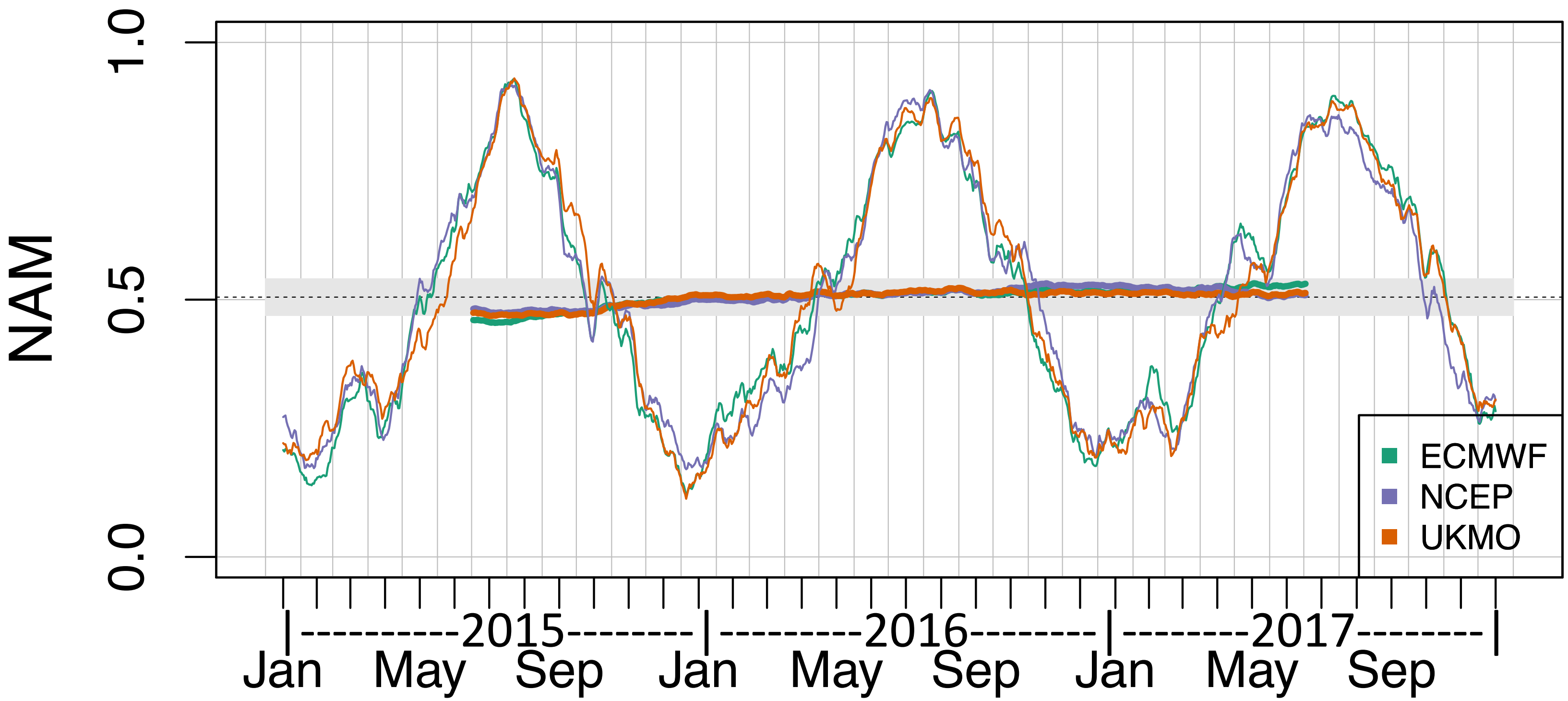


- ECDF normalization
  - NAM = (rank(PAM)-1/2)/n
  - Rank relative to ref. sample
- For minmax, and other normalizations
  - NAM = a PAM + b
  - a, b depend on ref. sample

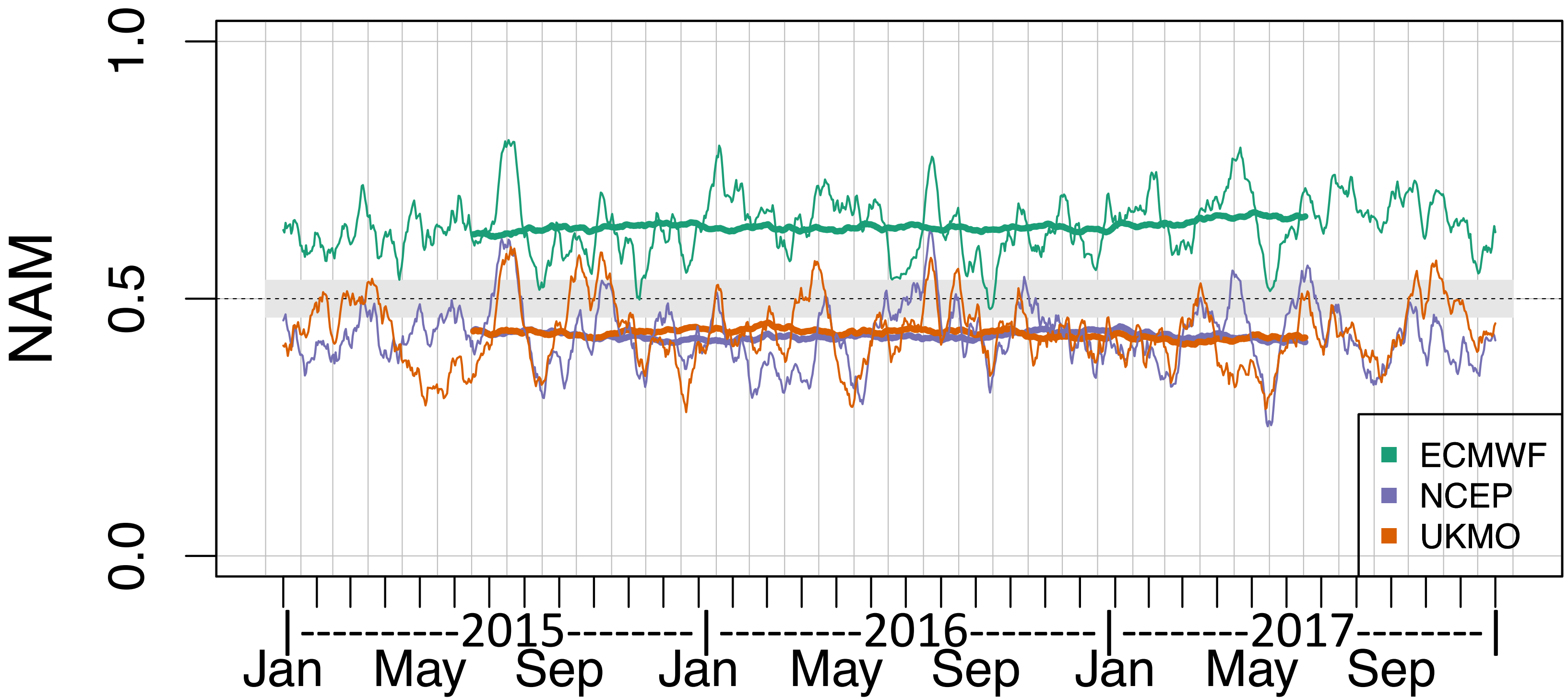
120-h 500-hPa NHX vector wind RMSE NAMs; All; MA(365) and MA(31)



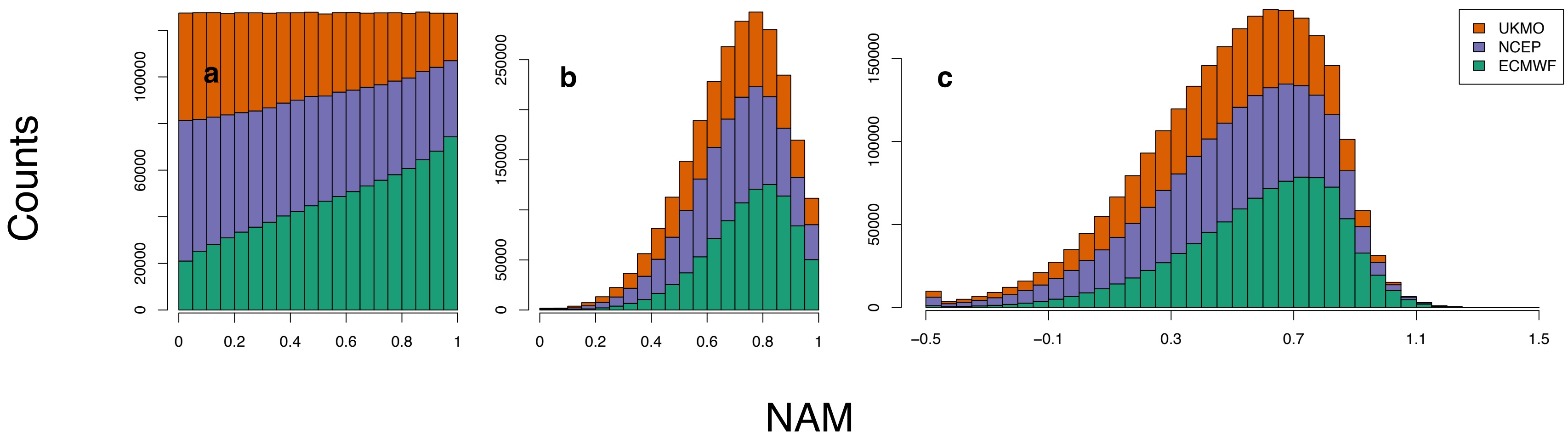
120-h 500-hPa NHX vector wind RMSE NAMs; ByCenter; MA(365) and MA(31)

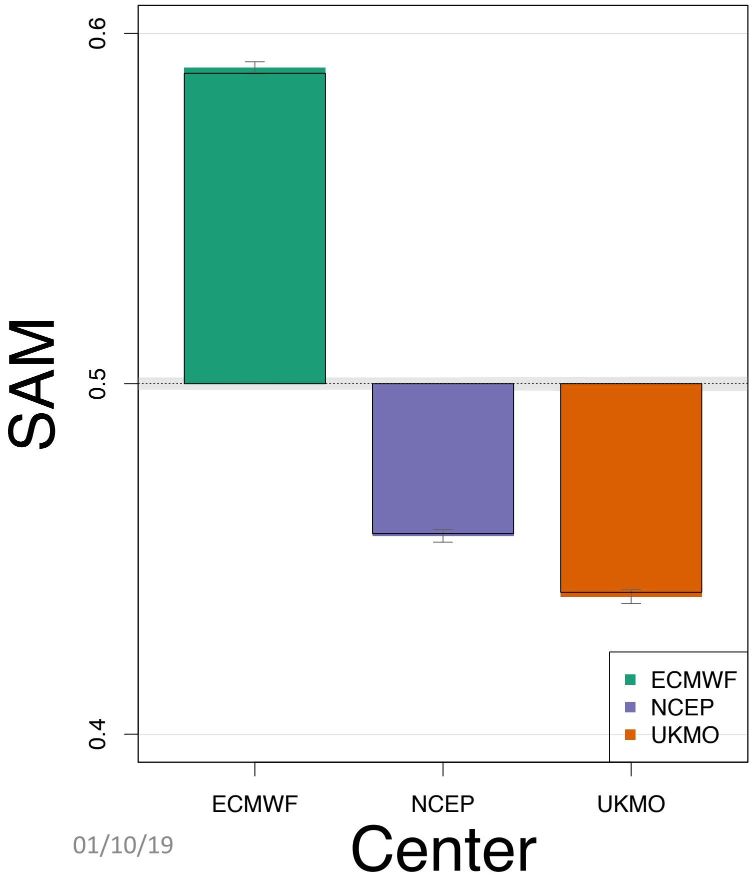


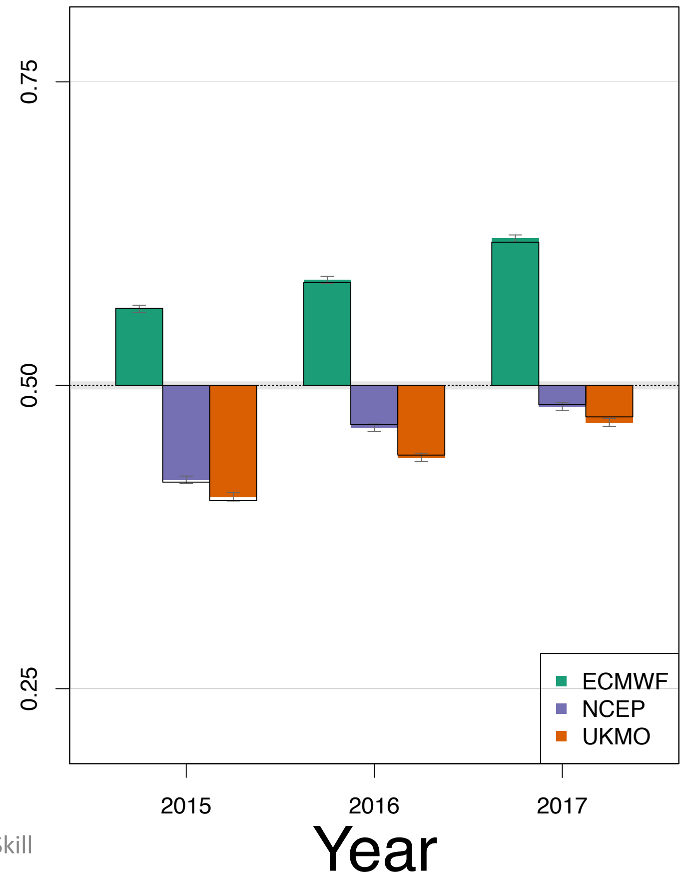
120-h 500-hPa NHX vector wind RMSE NAMs; ByMonth; MA(365) and MA(31)

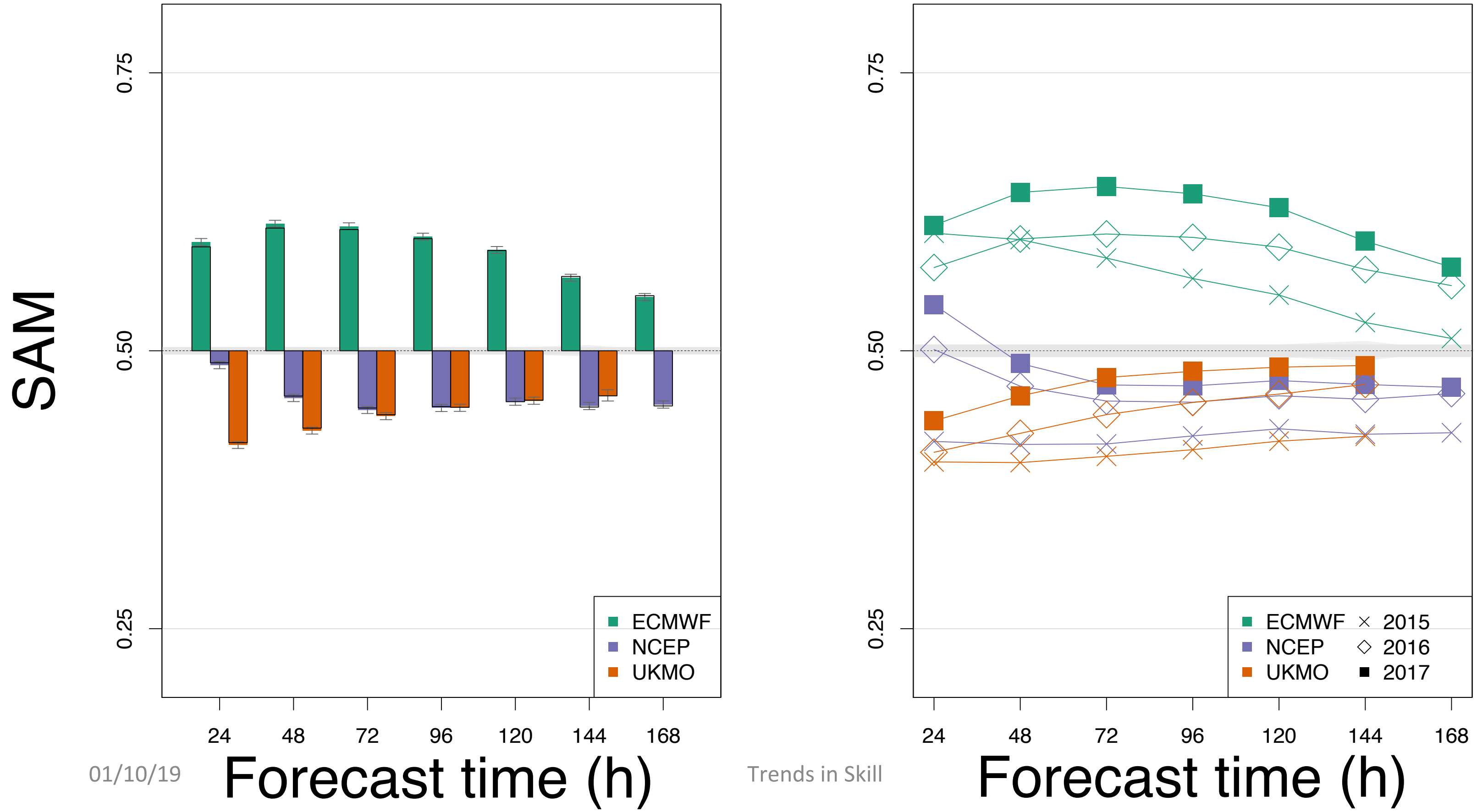


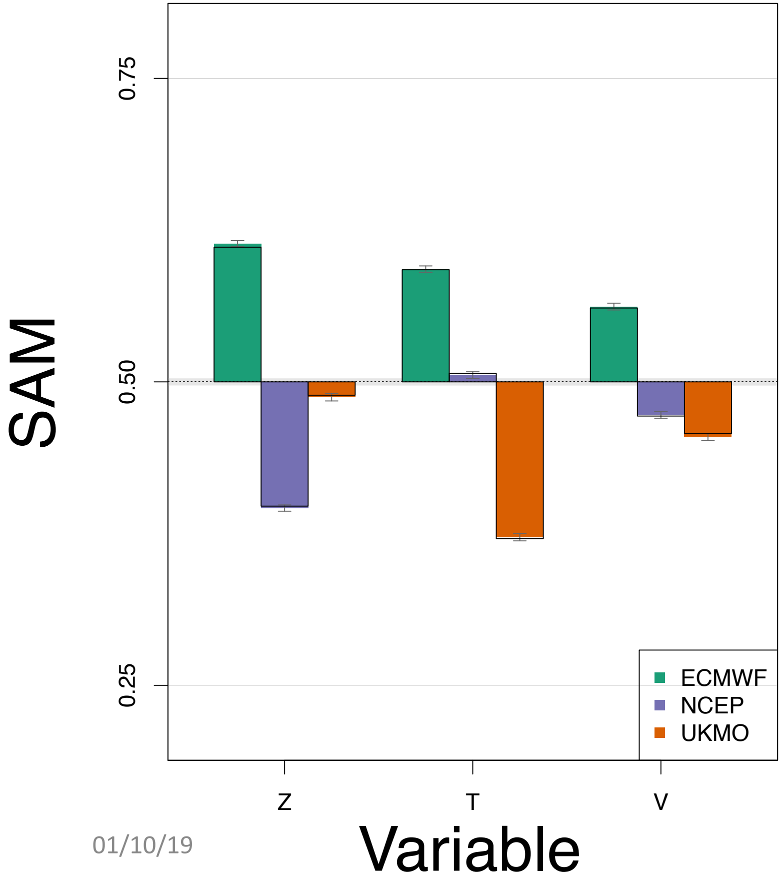
Histograms for all NAMs for (a) for ECDF, (b) for minmax, and (c) for rescaled-minmax normalizations

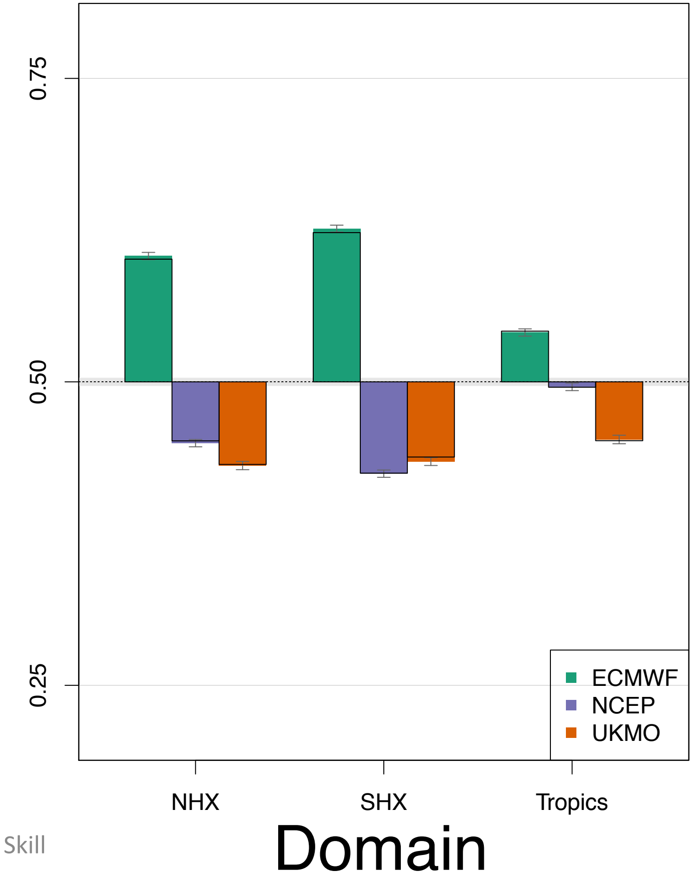


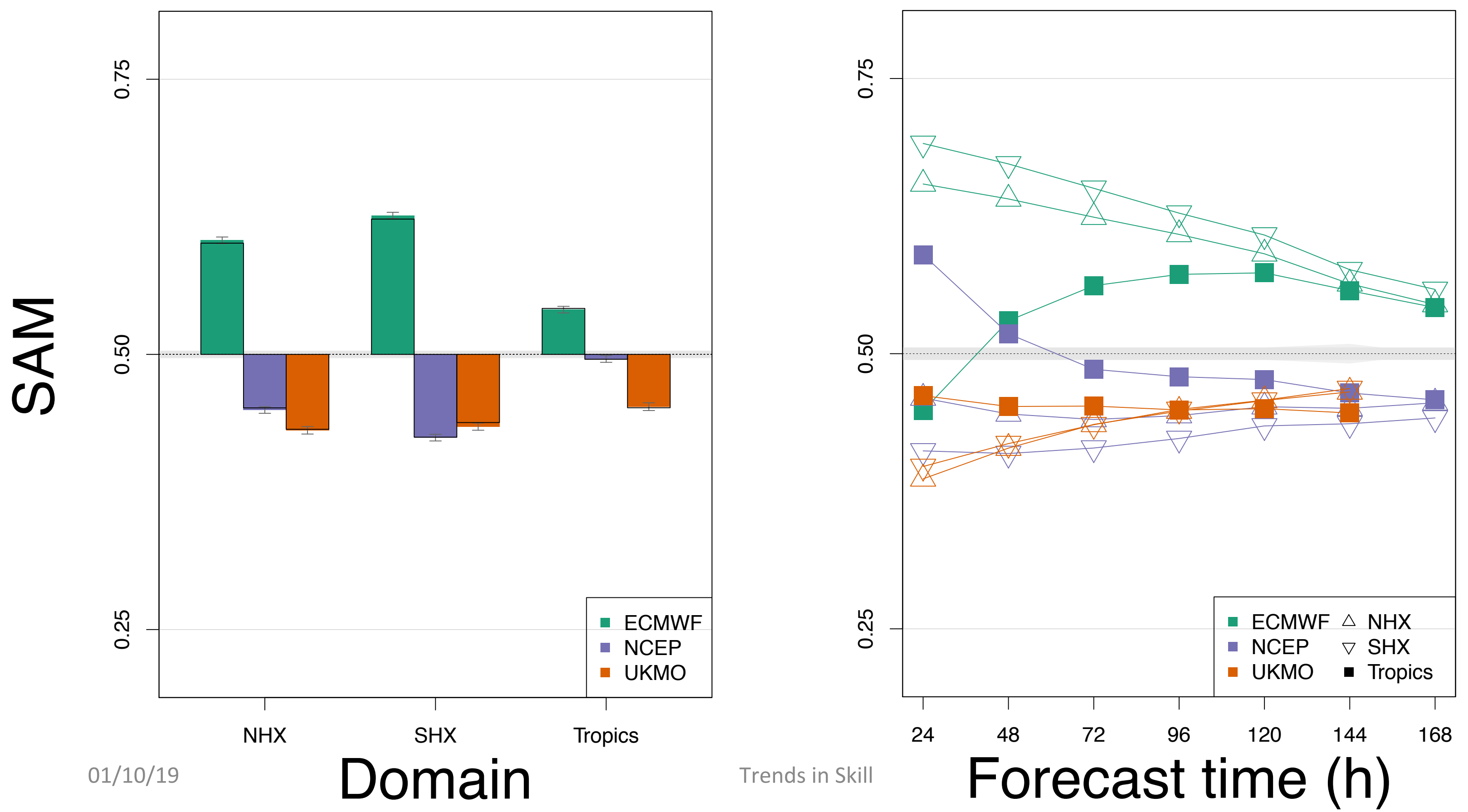

SAM
0.6
0.5
0.4
ECMWF
NCEP
UKMO
ECMWF
NCEP
UKMO
Center

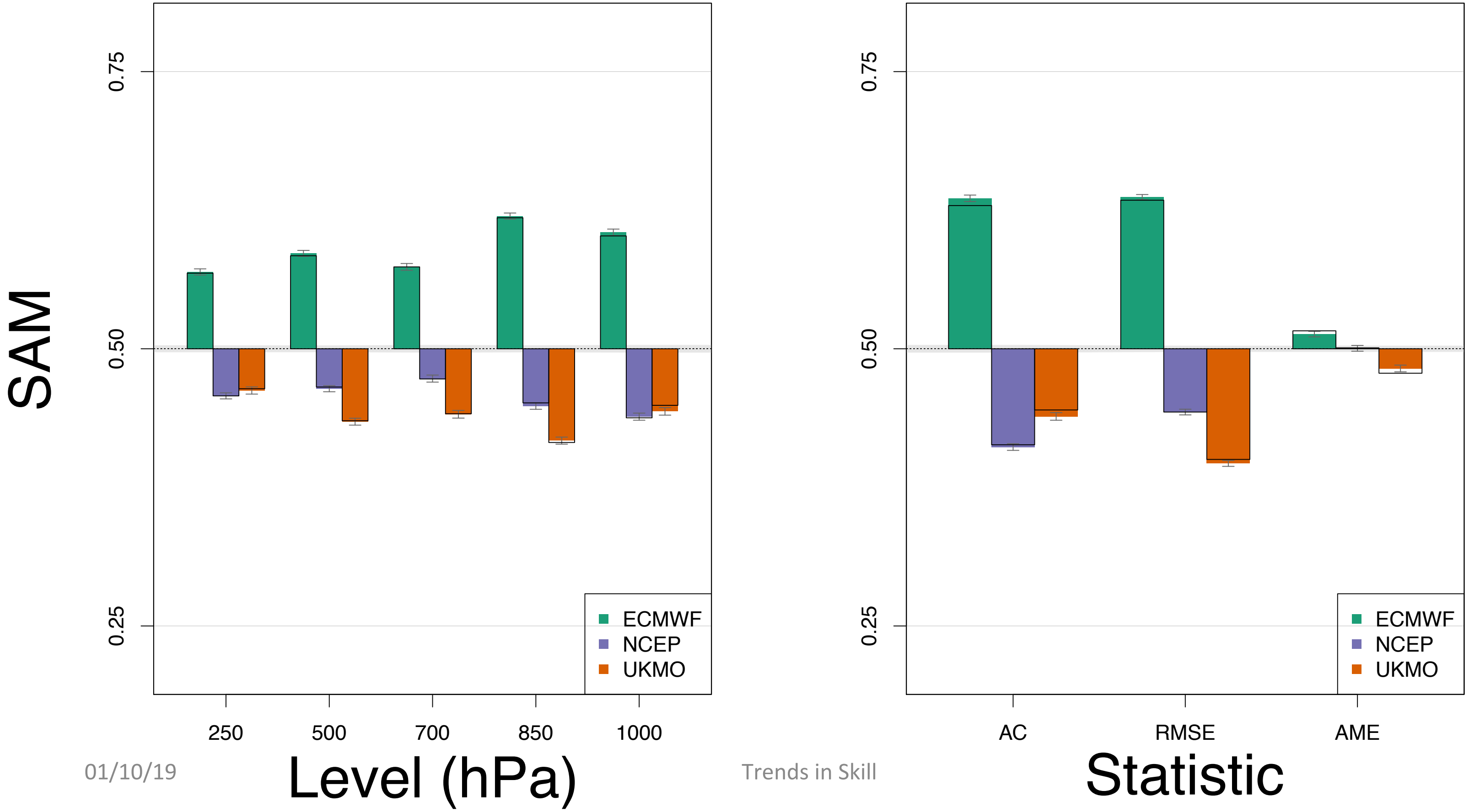

0.75
0.50
0.25
ECMWF
NCEP
UKMO
2015
2016
2017
Year


SAM
0.75
0.50
0.25
ECMWF
NCEP
UKMO
24
48
72
96
120
144
168
Forecast time (h)
0.75
0.50
0.25
ECMWF
NCEP
UKMO
2015
2016
2017
24
48
72
96
120
144
168
Forecast time (h)


SAM
0.75
0.50
0.25
ECMWF
NCEP
UKMO
Z
T
V
Variable

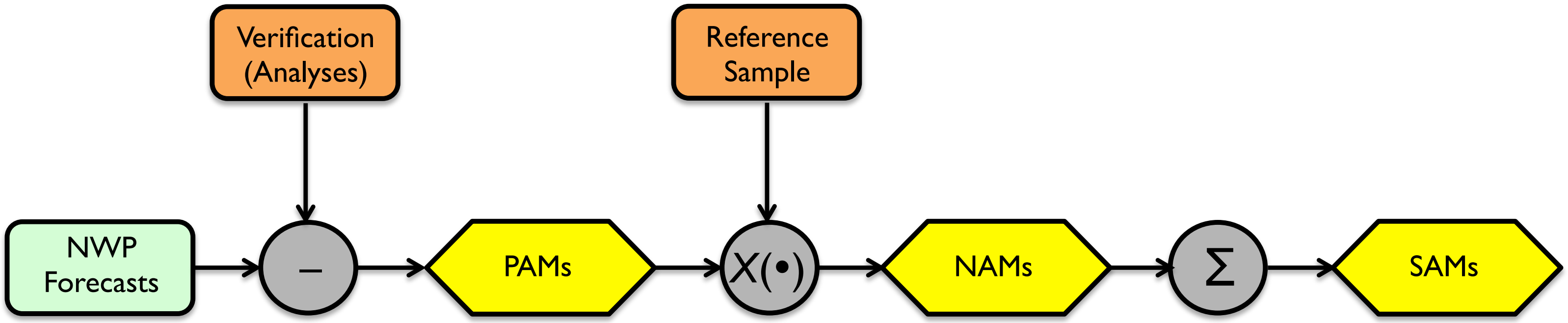

0.75
0.50
0.25
ECMWF
NCEP
UKMO
NHX
SHX
Tropics
Domain

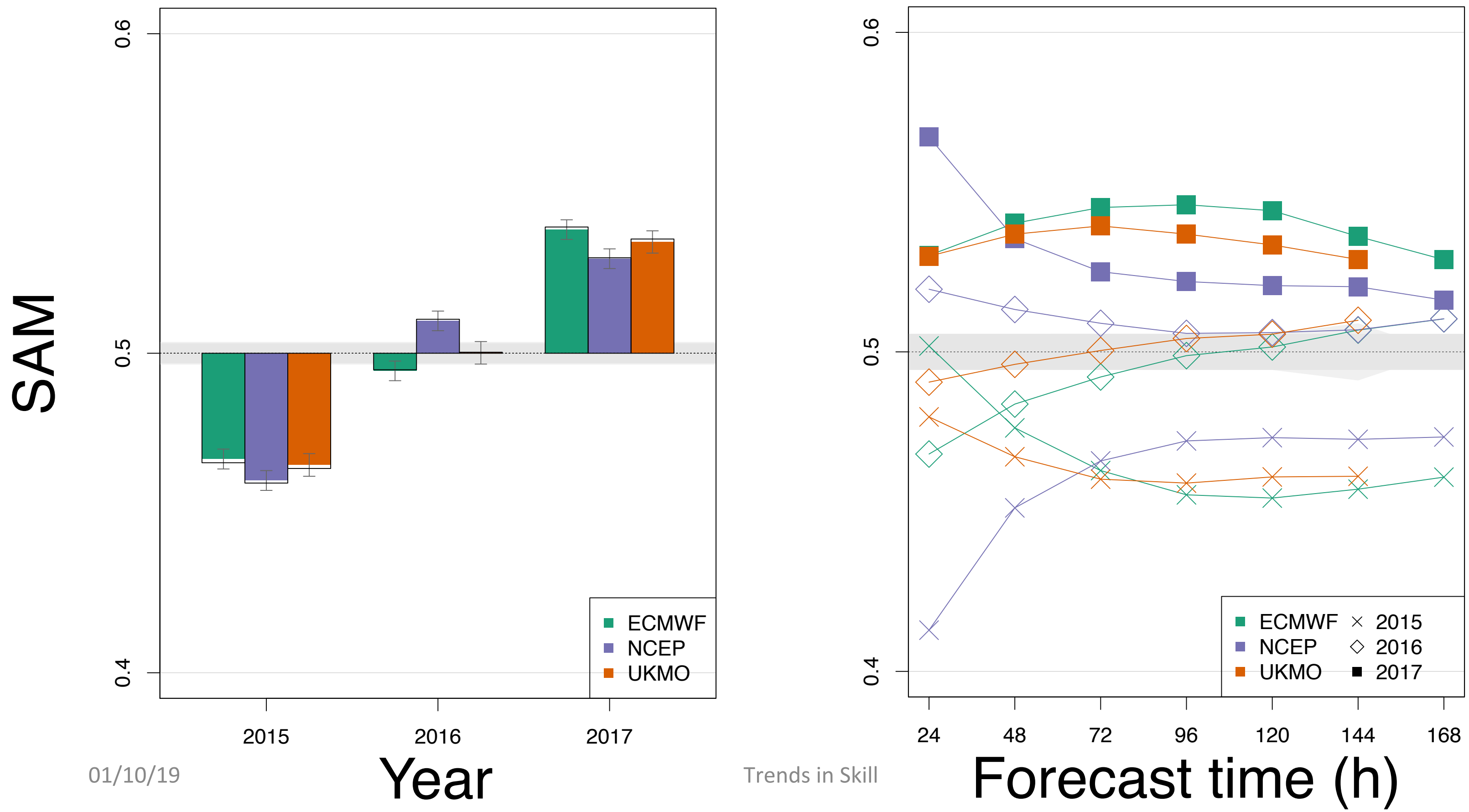

SAM
0.75
0.50
0.25
ECMWF
NCEP
UKMO
NHX
SHX
Tropics
Domain
0.75
0.50
0.25
ECMWF
NCEP
UKMO
NHX
SHX
Tropics
24
48
72
96
120
144
168
Forecast time (h)


SAM
0.75
0.50
0.25
ECMWF
NCEP
UKMO
250
500
700
850
1000
Level (hPa)
AC
RMSE
AME
Statistic

ByCenter normalization

- Reference sample for normalization
  - All :: (verification time, center)
  - ByCenter :: (verification time)



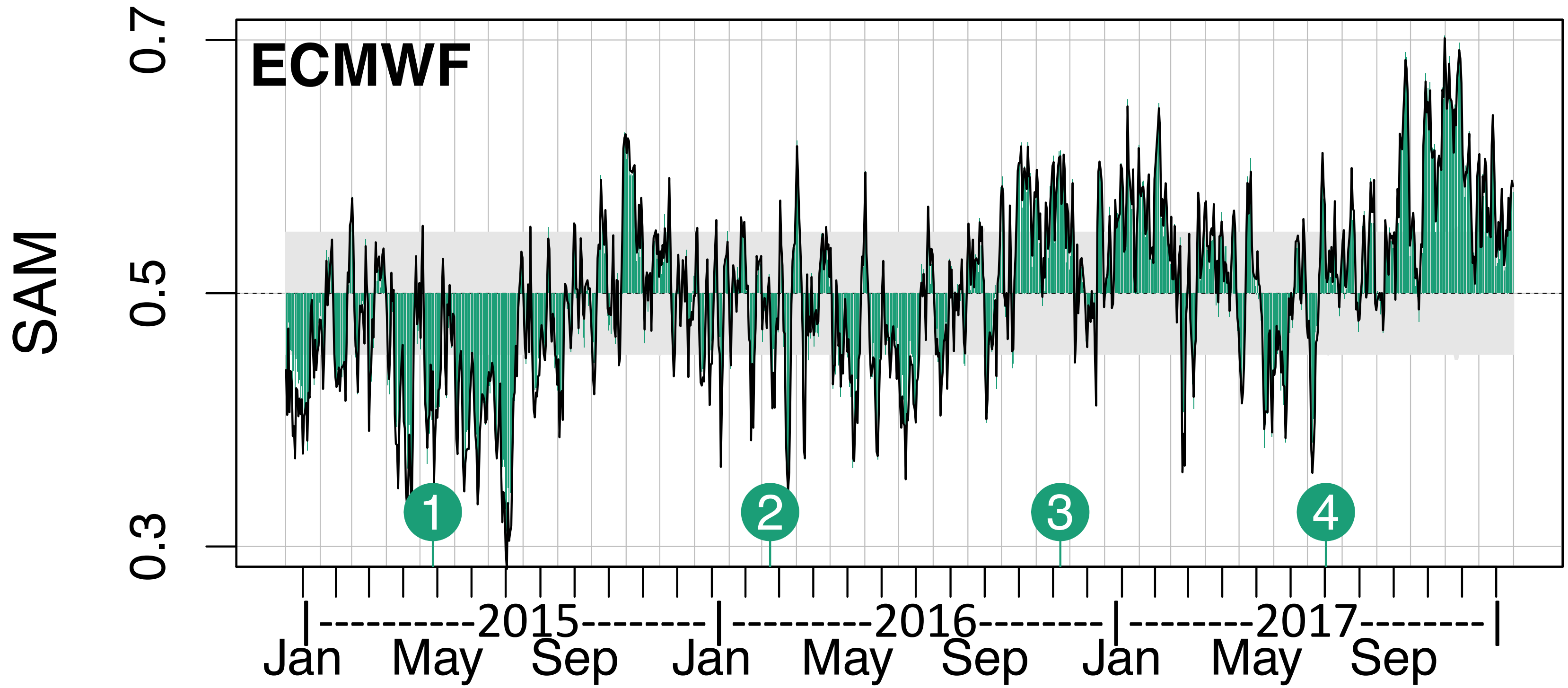

SAM
0.6
0.5
0.4
2015
2016
2017
Year
ECMWF
NCEP
UKMO
24
48
72
96
120
144
168
Forecast time (h)
ECMWF × 2015
NCEP ◇ 2016
UKMO ■ 2017

day-by-day

Upgrades 2015-2017

Center	i	Date	Upgrade	Delta
ECMWF	1	20150512	IFS Cycle 41r1	2.10
	2	20160308	IFS Cycle 41r2 (Cubic Octahedral 1280)	1.31
	3	20161122	IFS Cycle 43r1	2.58
	4	20170717	IFS Cycle 43r3	5.22
NCEP	1	20150114	TIN14-46 (T1534)	-4.12
	2	20160511	TIN16-11 (4DEnVar)	7.37
	3	20170719	SCN17-67 (NEMSIO)	0.81
UKMO	1	20161121	PS38 (satellite obs.)	4.75
	2	20170907	PS39 (10-km resolution)	2.82

day-by-day

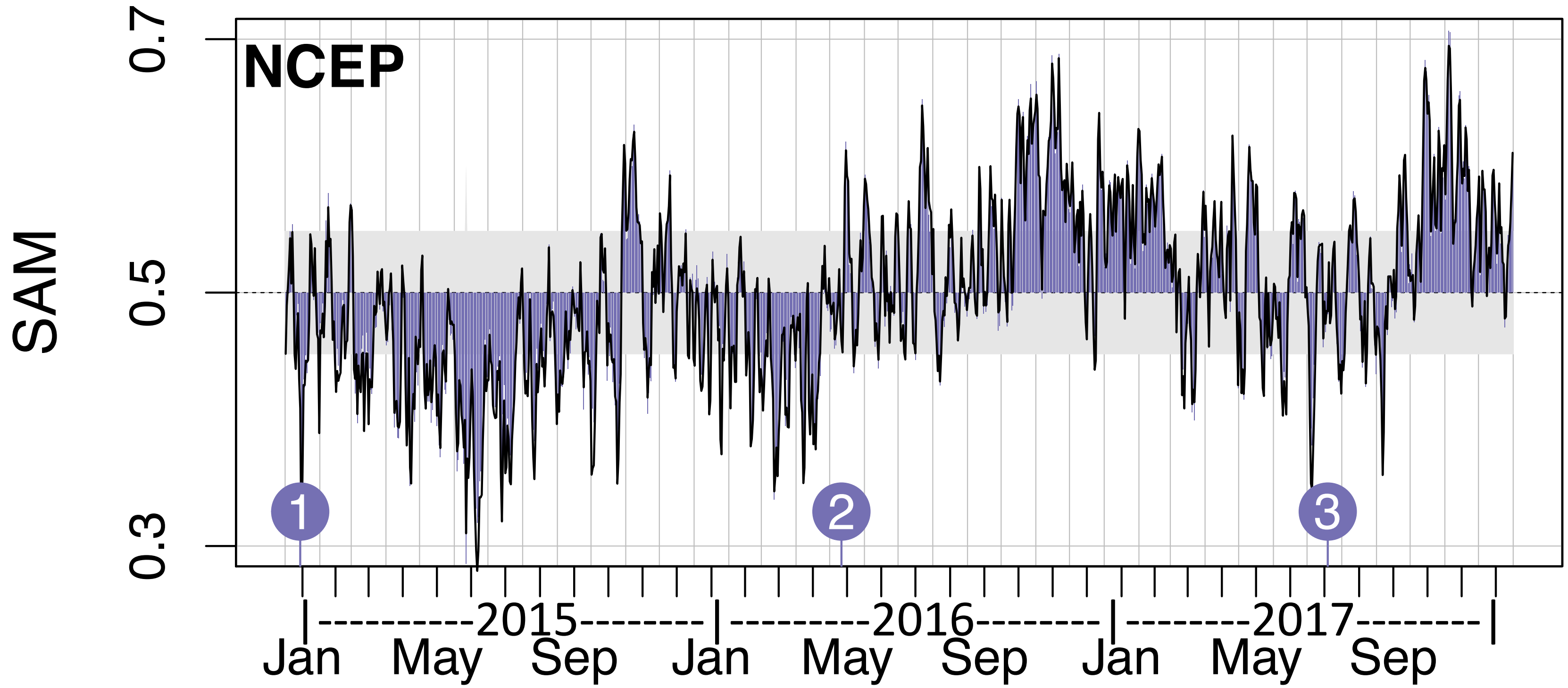

ECMWF
SAM
0.7
0.5
0.3
1
2
3
4
|---------2015--------|---------2016--------|--------2017--------|
Jan May Sep Jan May Sep Jan May Sep

day-by-day

Upgrades 2015-2017

Center	i	Date	Upgrade	Delta
ECMWF	1	20150512	IFS Cycle 41r1	2.10
	2	20160308	IFS Cycle 41r2 (Cubic Octahedral 1280)	1.31
	3	20161122	IFS Cycle 43r1	2.58
	4	20170717	IFS Cycle 43r3	5.22
NCEP	1	20150114	TIN14-46 (T1534)	-4.12
	2	20160511	TIN16-11 (4DEnVar)	7.37
	3	20170719	SCN17-67 (NEMSIO)	0.81
UKMO	1	20161121	PS38 (satellite obs.)	4.75
	2	20170907	PS39 (10-km resolution)	2.82

day-by-day

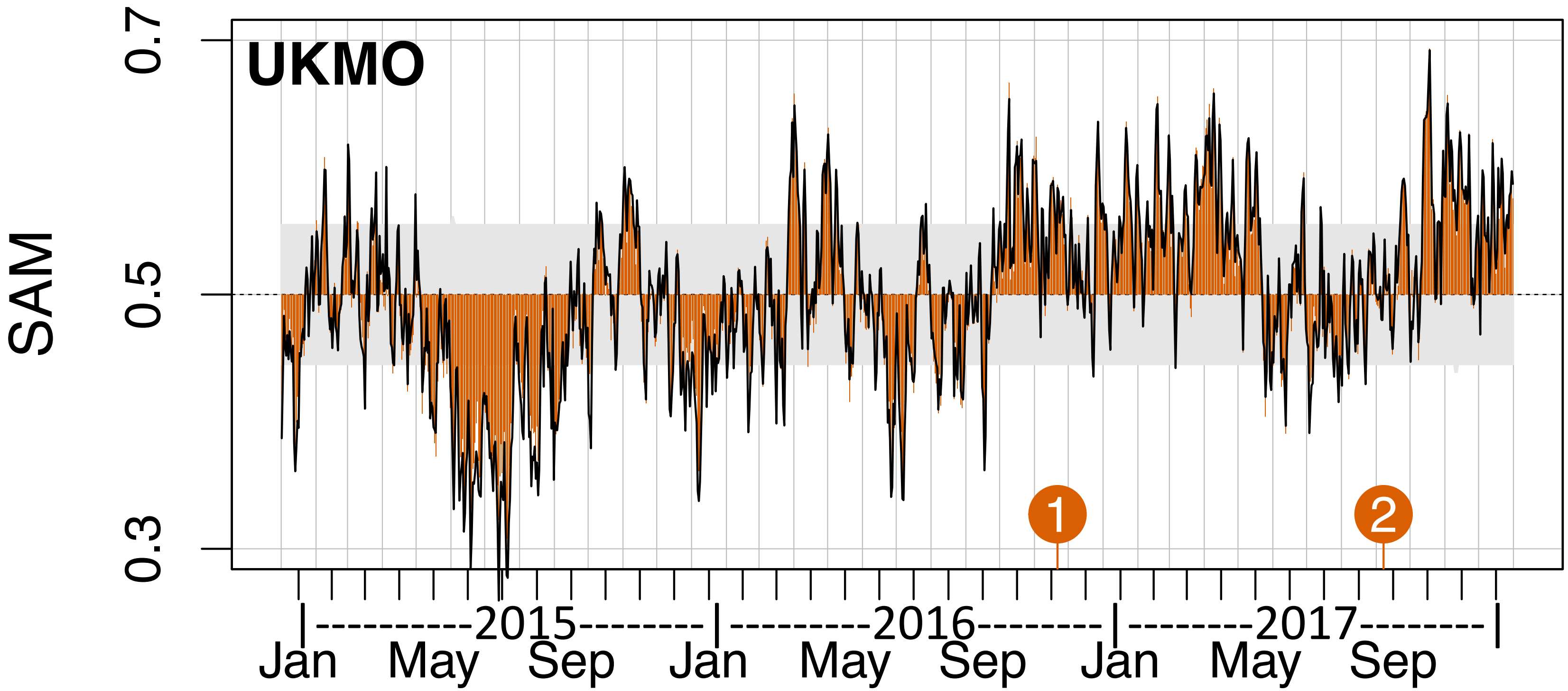

NCEP
SAM
0.7
0.5
0.3
1
2
3
|----------2015--------|---------2016--------|--------2017--------|
Jan May Sep Jan May Sep Jan May Sep

day-by-day

Upgrades 2015-2017

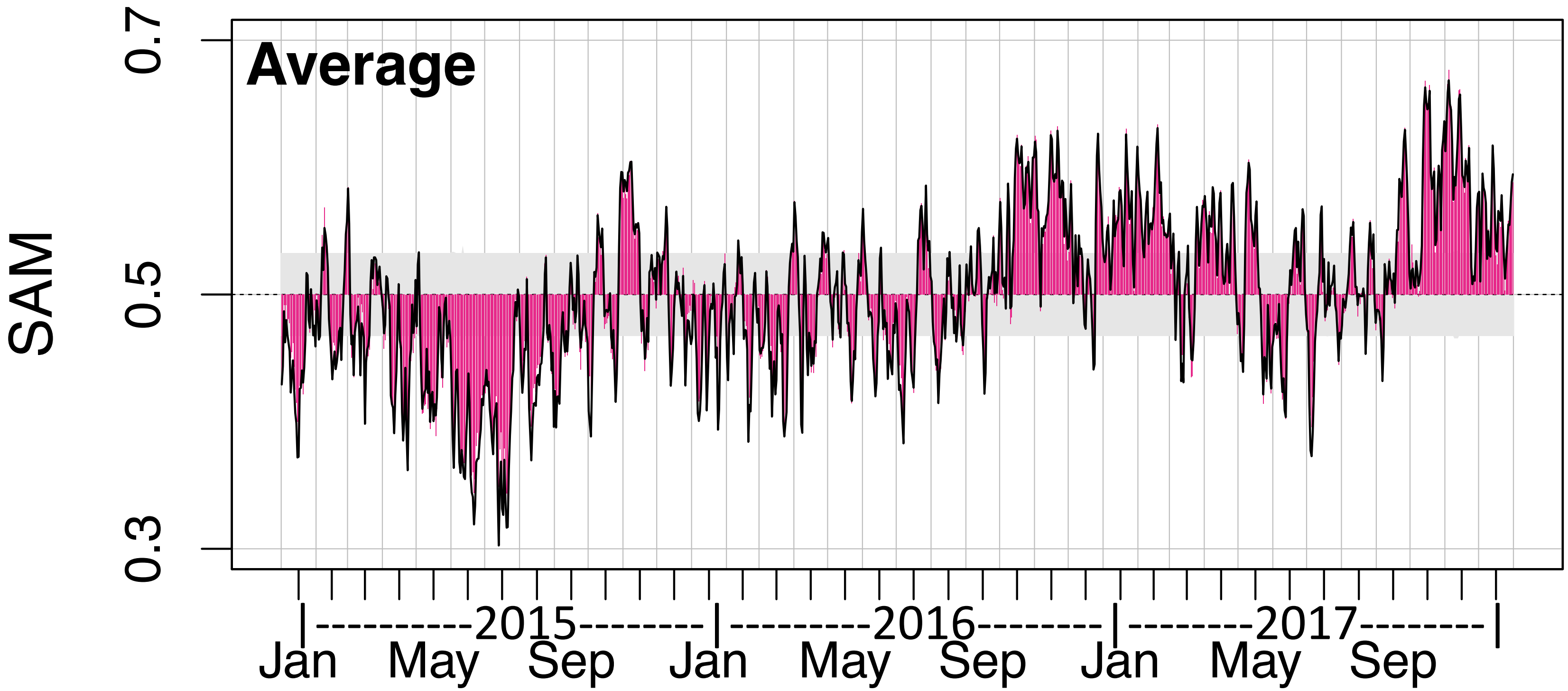
Center	i	Date	Upgrade	Delta
ECMWF	1	20150512	IFS Cycle 41r1	2.10
	2	20160308	IFS Cycle 41r2 (Cubic Octahedral 1280)	1.31
	3	20161122	IFS Cycle 43r1	2.58
	4	20170717	IFS Cycle 43r3	5.22
NCEP	1	20150114	TIN14-46 (T1534)	-4.12
	2	20160511	TIN16-11 (4DEnVar)	7.37
	3	20170719	SCN17-67 (NEMSIO)	0.81
UKMO	1	20161121	PS38 (satellite obs.)	4.75
	2	20170907	PS39 (10-km resolution)	2.82

day-by-day

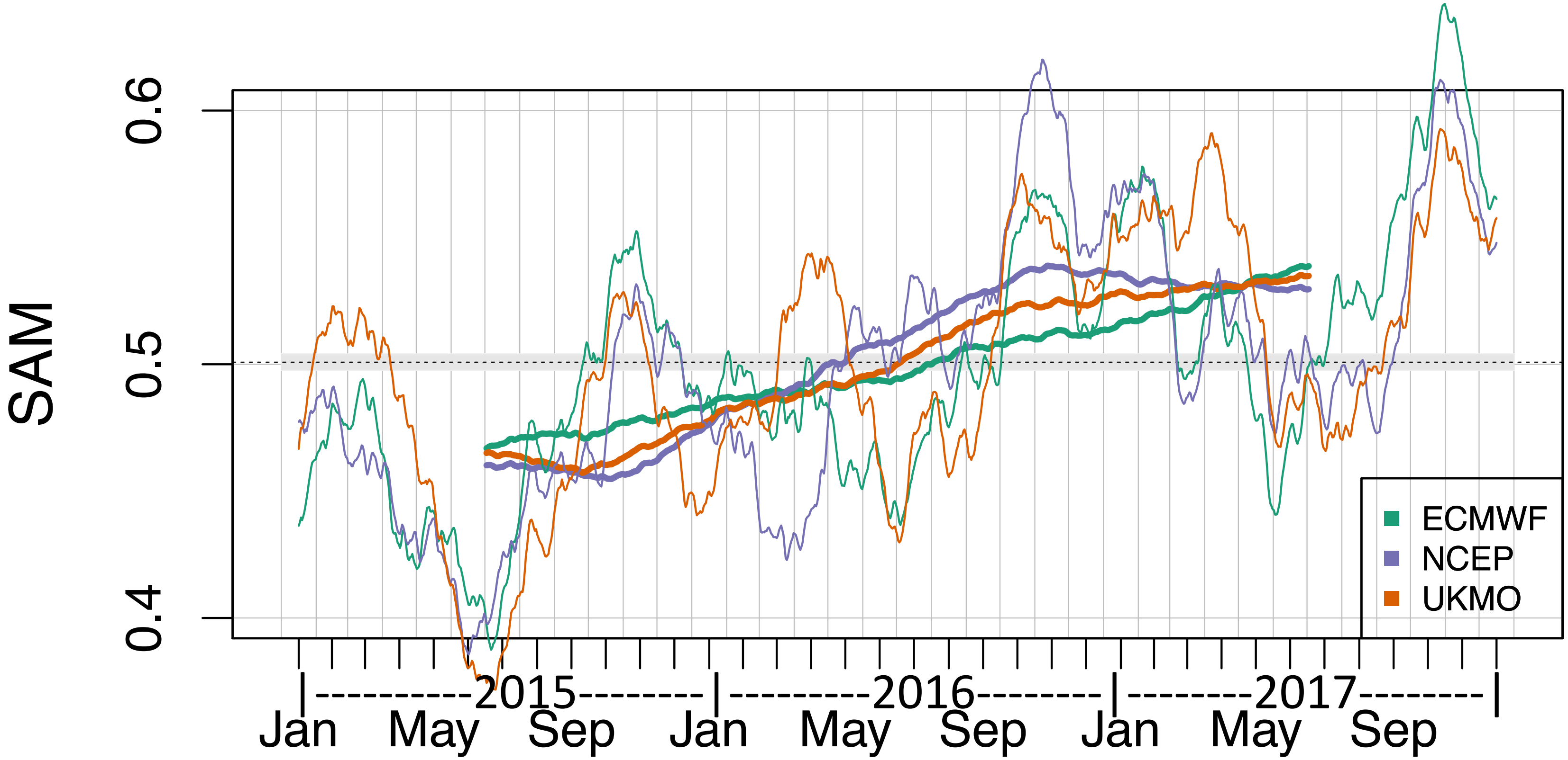


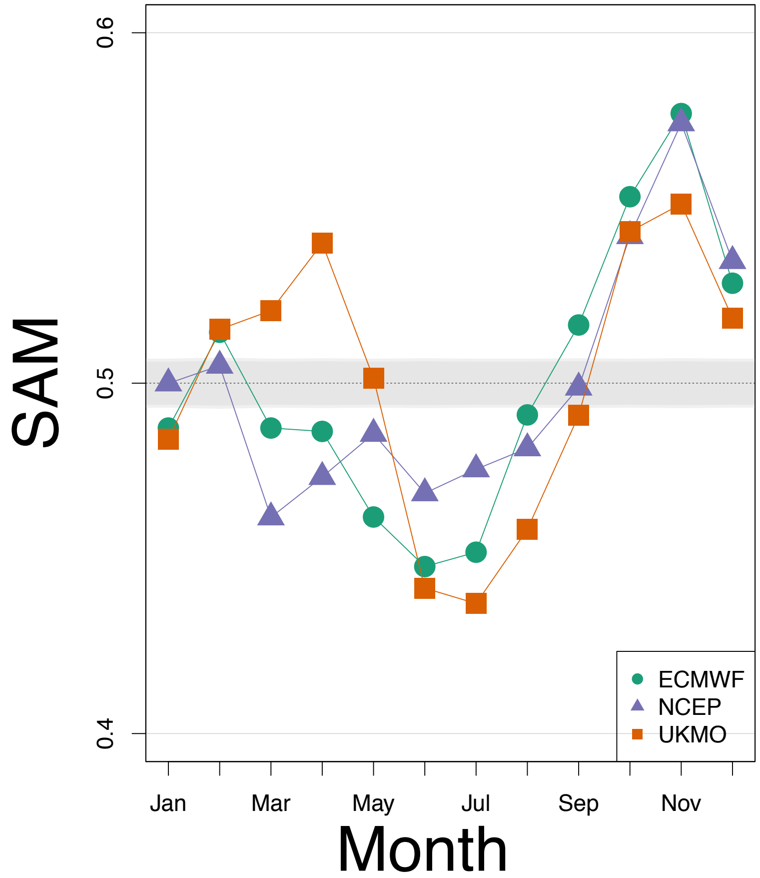
UKMO
SAM
0.7
0.5
0.3
1
2
|----------2015---------|---------2016---------|---------2017---------|
Jan May Sep Jan May Sep Jan May Sep

day-by-day


Average
SAM
0.7
0.5
0.3
|----------2015---------|---------2016---------|--------2017---------|
Jan
May
Sep
Jan
May
Sep
Jan
May
Sep

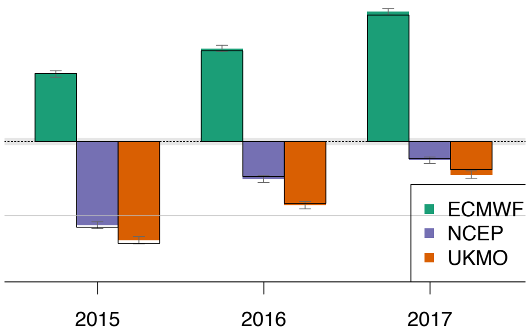
MA(365) and MA(31)




SAM
0.4
0.5
0.6
Jan
Mar
May
Jul
Sep
Nov
Month
ECMWF
NCEP
UKMO

Summary

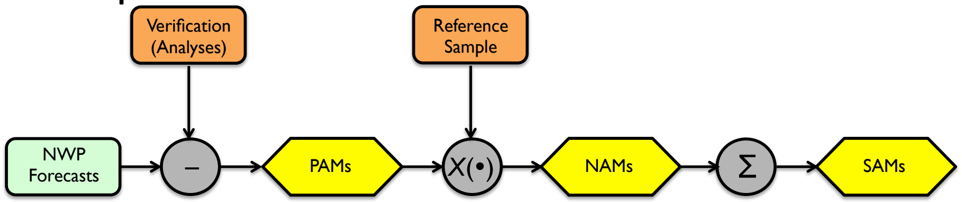
- All three centers improve over the three year period. NCEP short-term forecast skill substantially increases during the period.
- SAMs indicate that in terms of forecast skill ECMWF is better than NCEP, which is better than but approximately the same as UKMO.



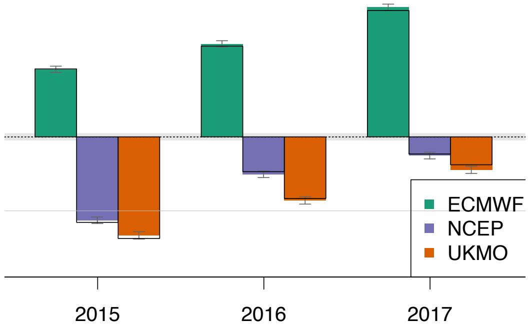
- However, the observed impacts are within the context of slowly improving forecast skill for operational global NWP as compared to earlier years.
- The use of SAMs improves the signal to noise ratio and clear improvements in SAM are related to the ECMWF July 2017 upgrade to IFS Cycle 43r3, the NCEP May 2016 replacement of the 3DEnVar with the 4DEnVar, and the UKMO November 2016 (PS38) introduction of improved use of satellite observations.

Concluding remarks

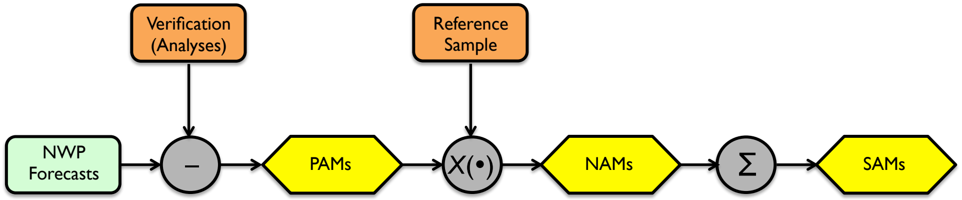
- We often focus on a few key PAMs, but this may ignore other important aspects of forecast skill. The use of SAMs increases statistical significance and enables exploring different aspects of forecast skill.
- Clearly the systems lagging ECMWF can improve, and there is evidence from SAMs in addition to the 4DEnVar example that improvements in forecast and data assimilation systems are still leading to forecast skill improvements.
- In future work, it might be interesting to include other centers and to add PAMs for relative humidity and precipitation, forecast variables for which there is currently major room for improvement.



more…

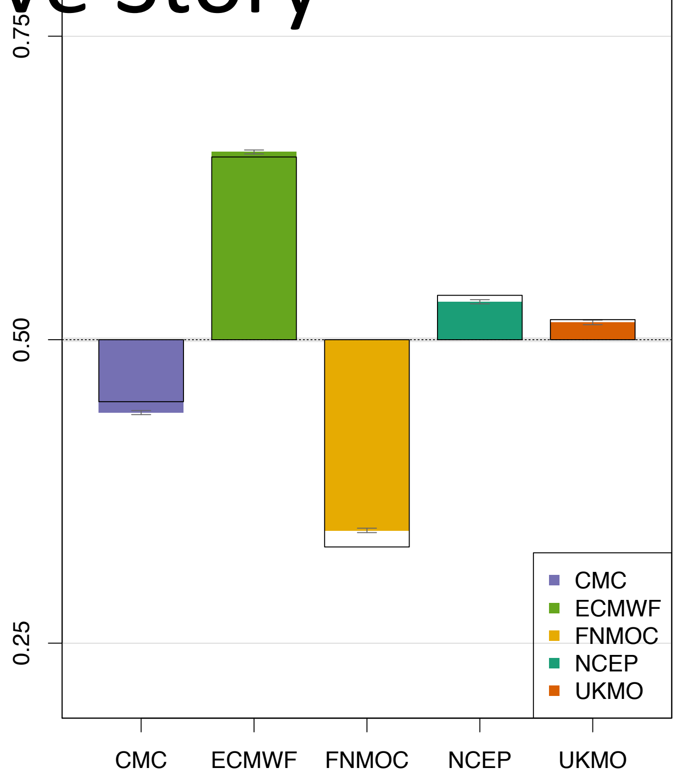


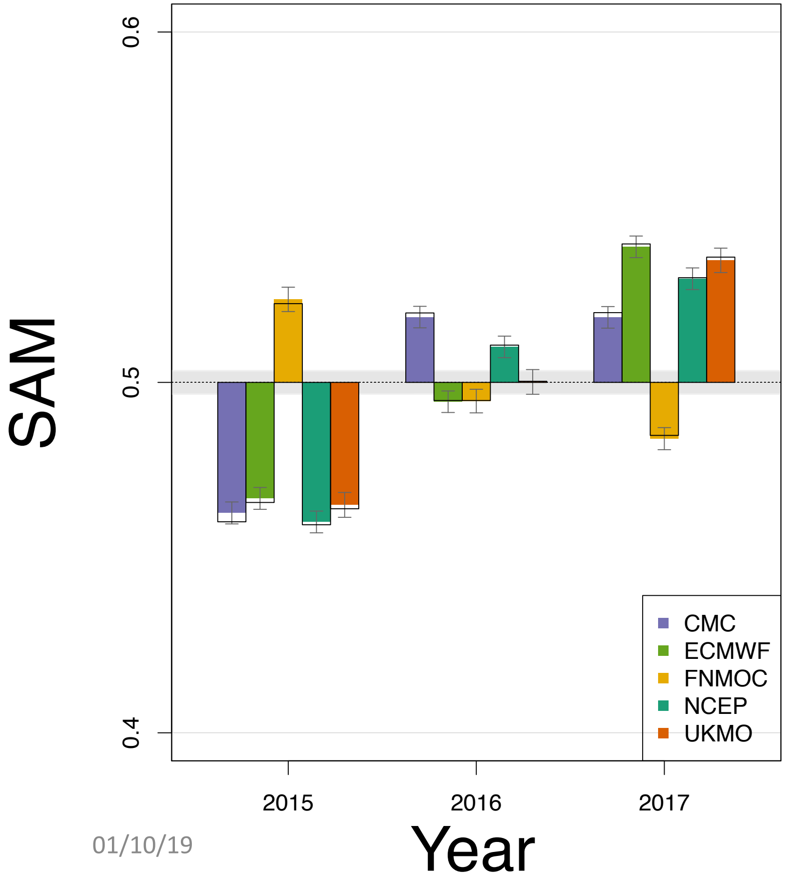
- email:
  - ross.n.hoffman@noaa.gov
- Dec 2018 WAF paper:
  - doi: 10.1175/WAF-D-18-0117.1
- AMS presentation:
  - https://ams.confex.com/ams/2019Annual/meetingapp.cgi/Paper/350739

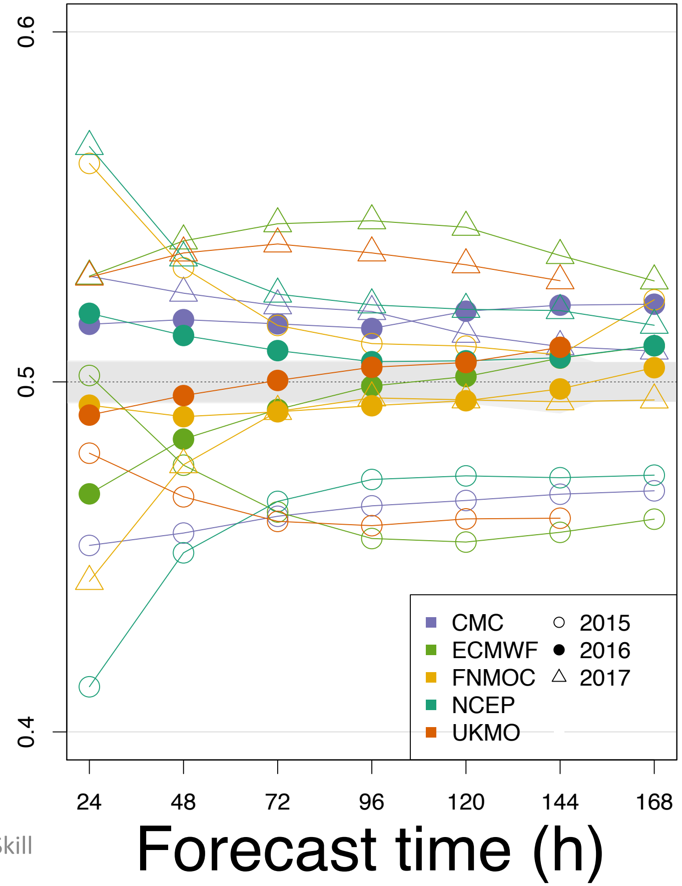


A Detective Story

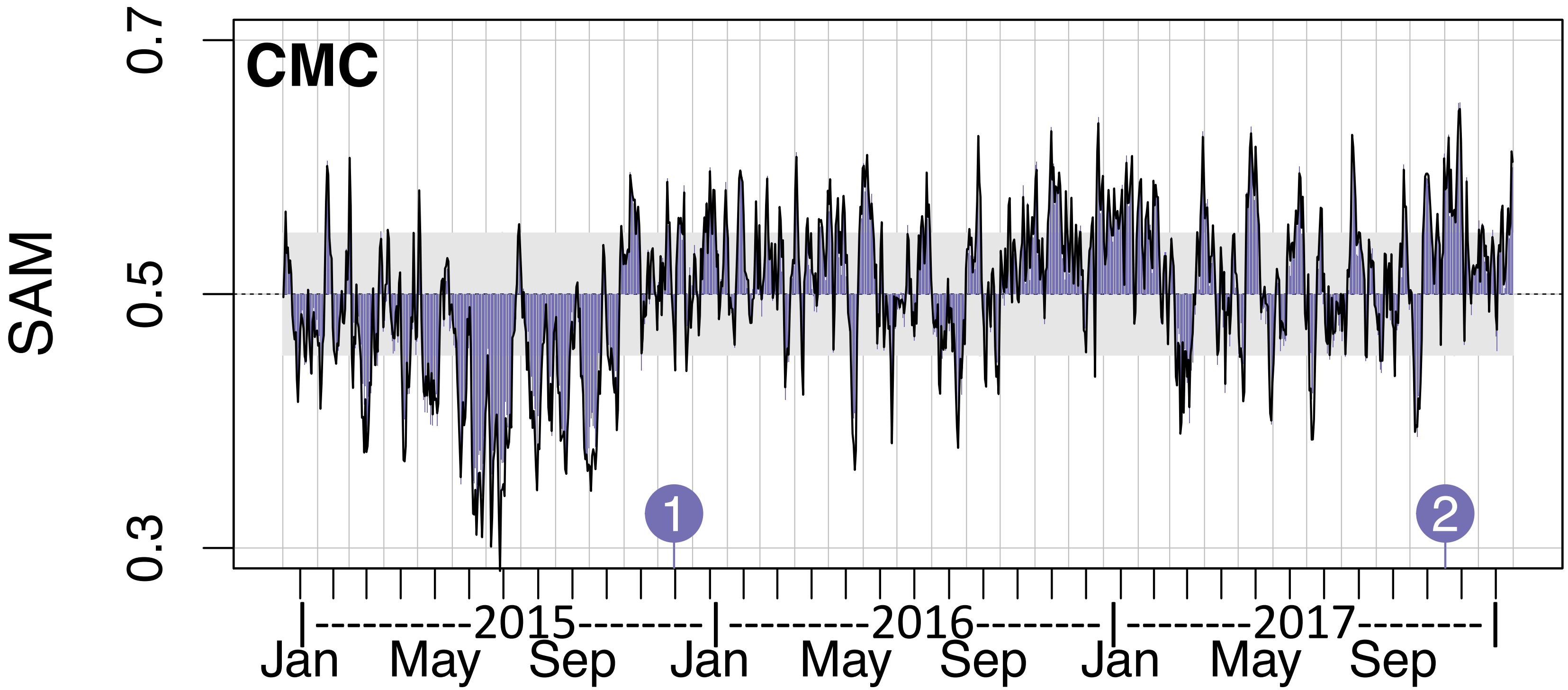
- We began with 5 centers: CMC, ECMWF, FNMOC, NCEP, UKMO
- We kept the 3 best, partly because FNMOC results did not make sense
- Colors have changed from the 3 centers case.




SAM
0.6
0.5
0.4
2015
2016
2017
Year
CMC
ECMWF
FNMOC
NCEP
UKMO

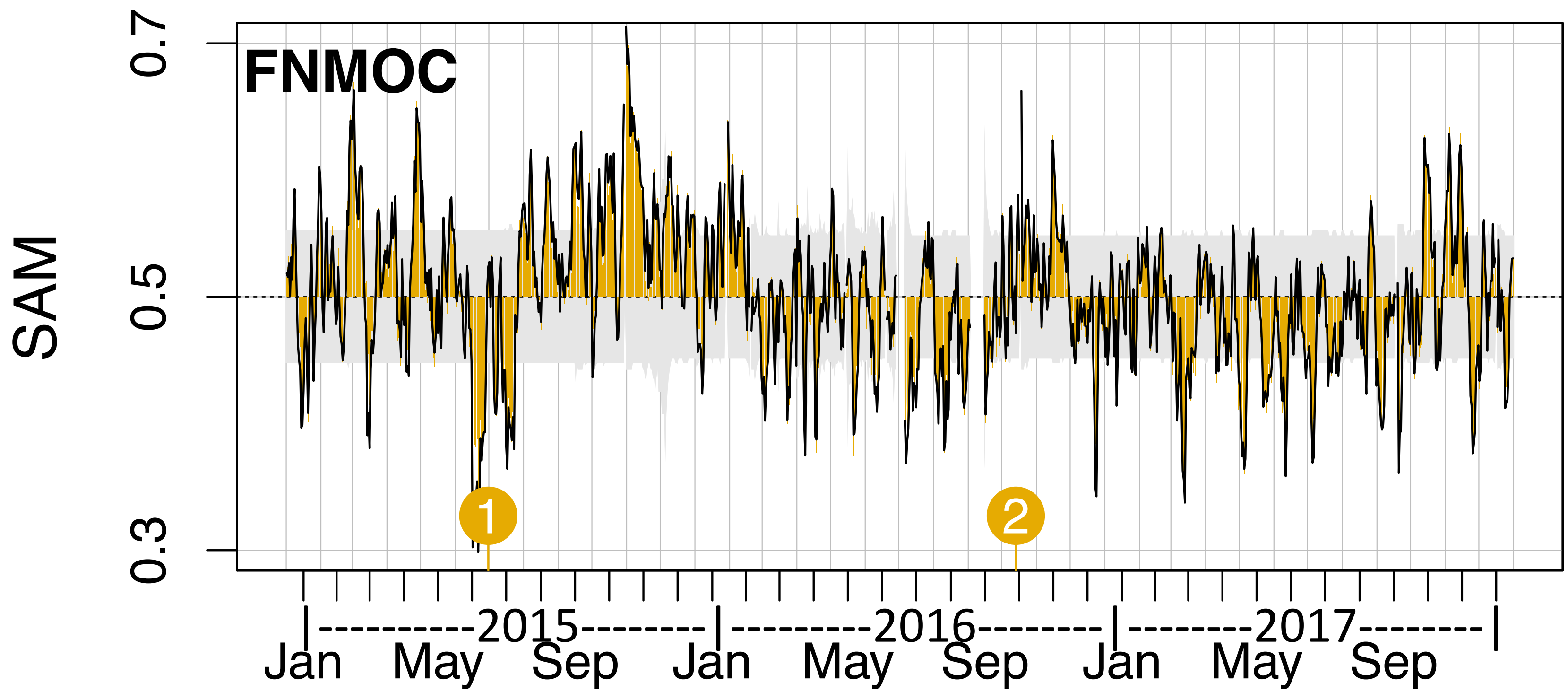

0.6
0.5
0.4
24
48
72
96
120
144
168
Forecast time (h)
CMC
ECMWF
FNMOC
NCEP
UKMO
2015
2016
2017

day-by-day



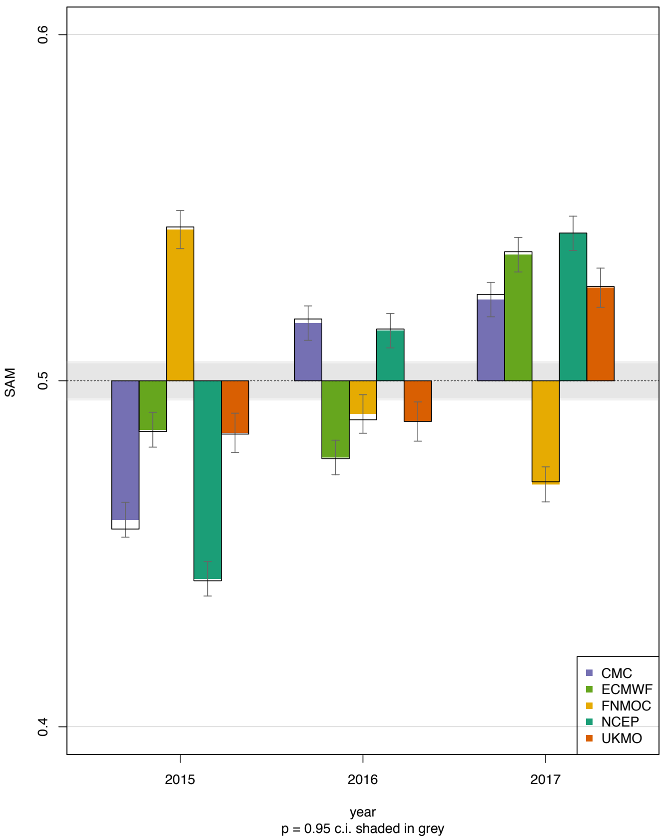
CMC
SAM
0.7
0.5
0.3
1
2
|----------2015--------|---------2016--------|--------2017--------|
Jan
May
Sep
Jan
May
Sep
Jan
May
Sep

day-by-day


FNMOC
SAM
0.7
0.5
0.3
1
2
|----------2015---------|---------2016---------|---------2017---------|
Jan
May
Sep
Jan
May
Sep
Jan
May
Sep

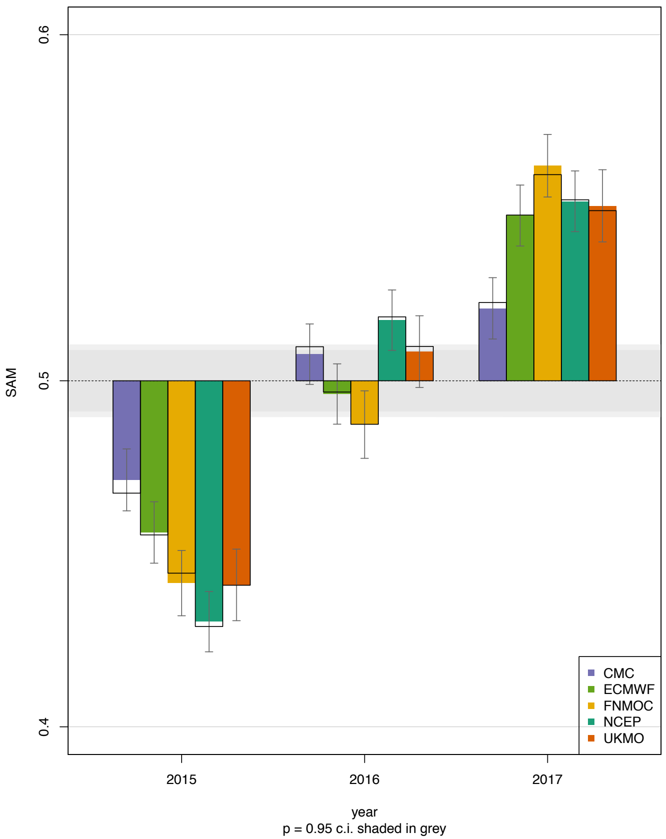
SAM

Wind (250, 500, 700, 850)



Year

Wind (1000)



Year

Why?

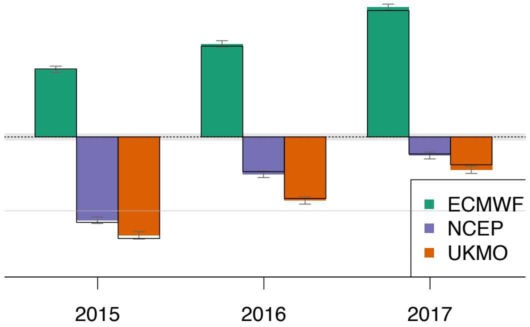
- In mid 2016, the NAVGEM grids shared with NCEP changed from 1 degree to 1/2 degree latitude-longitude.
- Because of the way the fields are filtered, this change makes it seem like NAVGEM forecast skill is degrading in our assessments using the VSDB statistics.
  - Most of the NAVGEM fields used in our assessments are filtered with the same 2d one-pass Shapiro smoother de-smoother applied in grid point space for both 1 and 1/2 degree fields.
  - As a result the 1/2 degree fields have more energy present in the inherently hard to predict smallest scales, resulting in an apparent drop in forecast skill.
  - During VSDB processing no further filtering is applied.

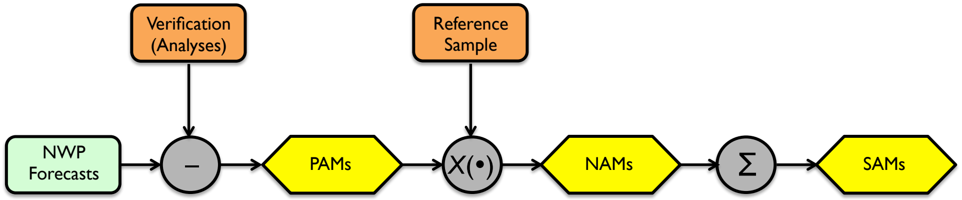
But what about low level winds?

- Filtering is applied to all geopotential heights, all temperatures, and all winds above 900 hPa.
  - Thus in our study, only 1000 hPa winds were not affected by this change.
- Thanks to Elizabeth Satterfield/NRL-Monterey and Randal Pauley/FNMOC for help in unraveling this puzzle.

more...

- email:
  - ross.n.hoffman@noaa.gov
- Dec 2018 WAF paper:
  - doi: 10.1175/WAF-D-18-0117.1
- AMS presentation:
  - https://ams.confex.com/ams/2019Annual/meetingapp.cgi/Paper/350739

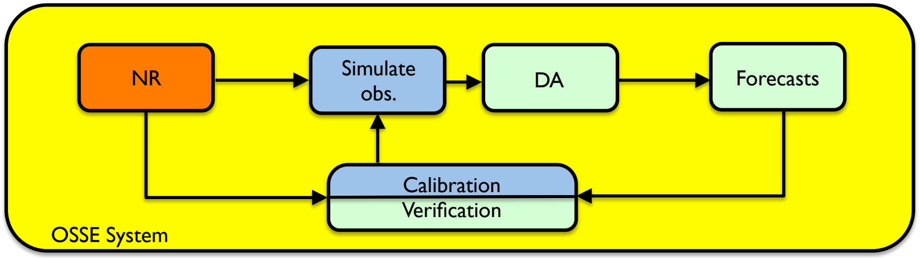




Application to impact experiments

- The origin of the SAM work was to summarize OSE results for a data gap imact study.
- We repeated the impact study in simulation (OSSE mode) to validate our OSSE system (CGOP).
- Ideally each OSSE component must be realistic, however these OSSEs used
  - Lower resolution GDAS/GFS
  - No observation errors

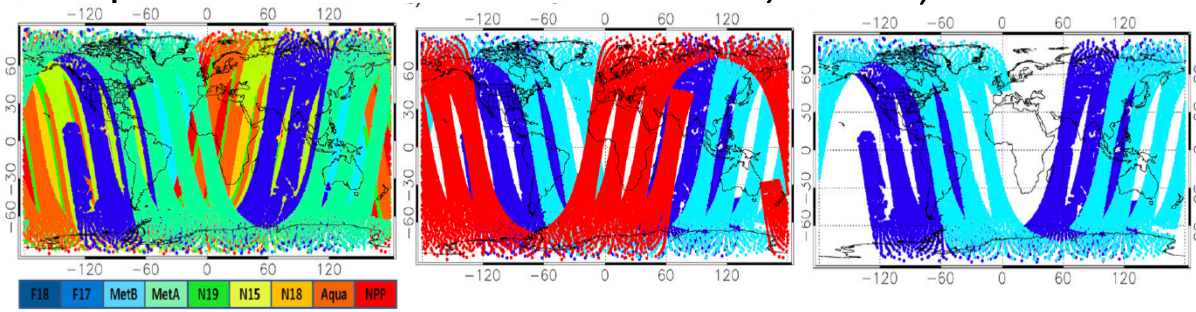
Observing System Simulation Experiments



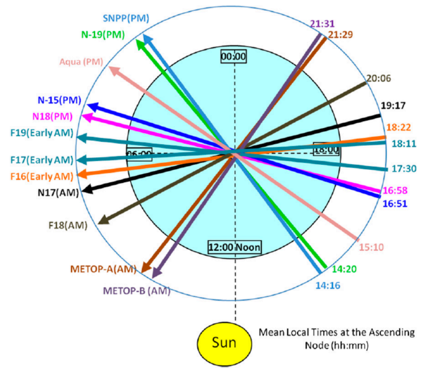
- OSSEs have been used since the 1950s to
  - Evaluate observing systems in terms of accuracy and coverage (e.g., in planning FGGE)
  - Guide decision makers to allocate resources to mitigate costs and lead time in reality
  - Conduct trade studies of instruments and systems, and
  - Design and test new DA methods

Experimental Setup: Data Gap Scenario

- Inter-comparison: OSSE vs OSE for Control, 3Polar, 2Polar



Period	7 July to 7 August, 2014
Obs system config.	2014: all conventional + satellite data gap scenarios
	OSSE only: perfect obs
DAS	3DEnVar with T670/T254 n=80 ensemble
Forecast	0000 UTC daily up to 168 h
Verification	OSSE/OSE: own cntrl analysis

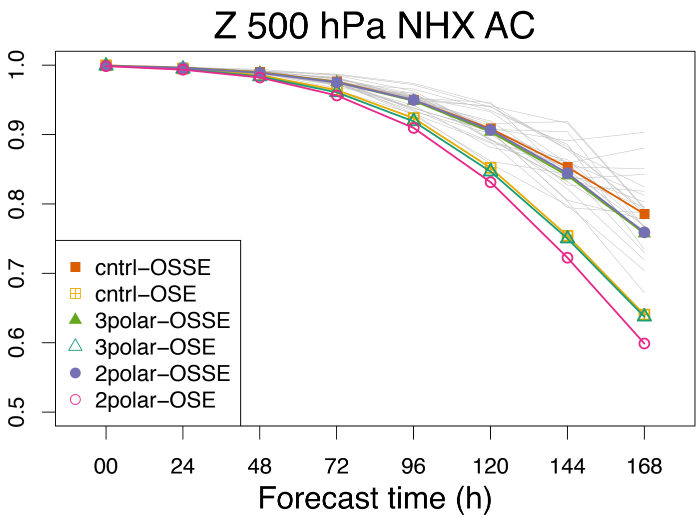


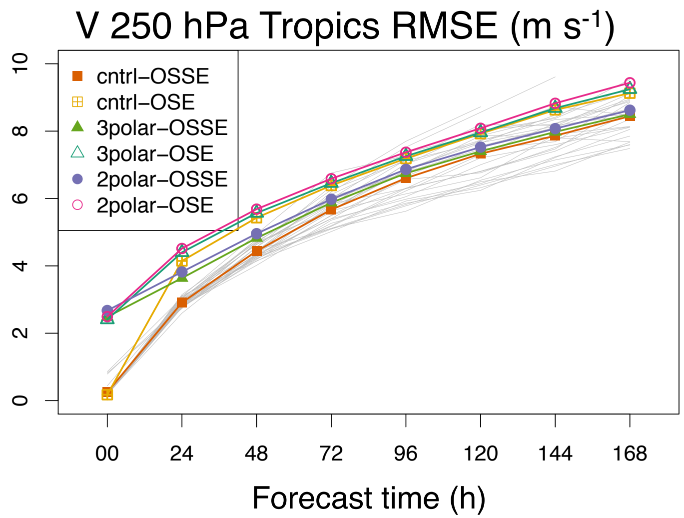
[Original OSE work by Boukabara et al (2016b)]

OSSE system validation/calibration

- Before apply an OSSE system to a new proposed sensor, we want to check if the results for a data denial experiment match
- If not (and often OSSE results are overly optimistic) we must calibrate our OSSE system by
  - Tuning the observation errors; and/or
  - Changing some parameterizations in the forecast model; or
  - Adjusting the OSSE results after the fact
- For the simulation experiments with "perfect" observations, the OSSE was overly optimistic, but not in terms of SAMs!

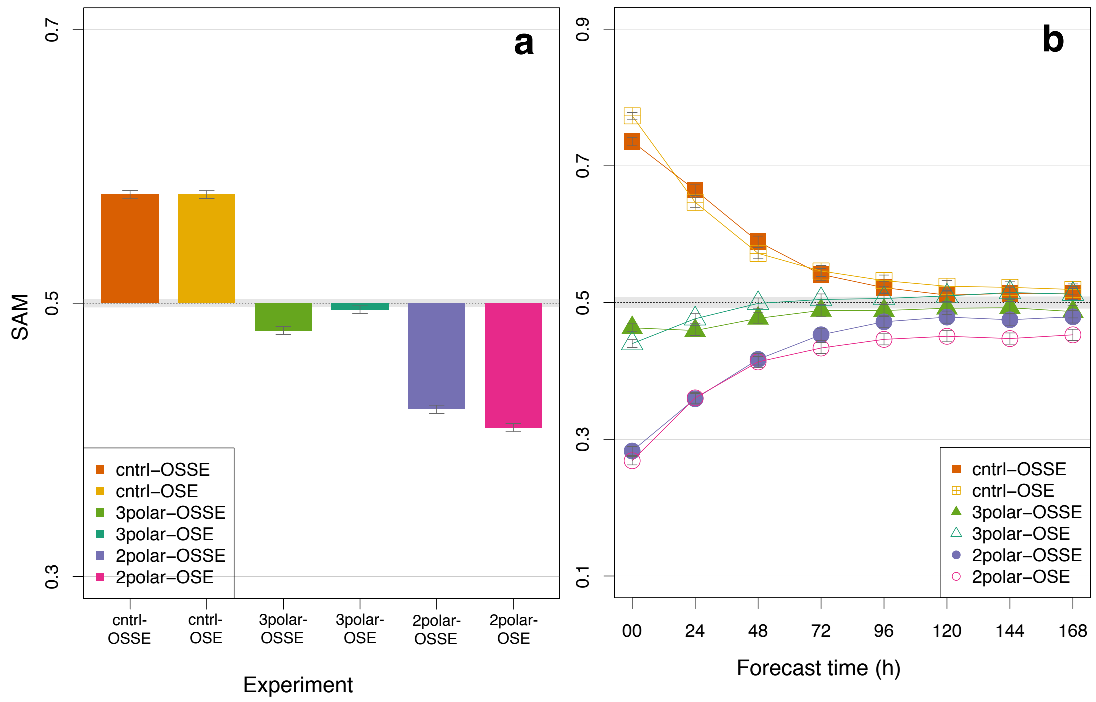
Results: Forecast Skill (PAM - AC & RMSE)



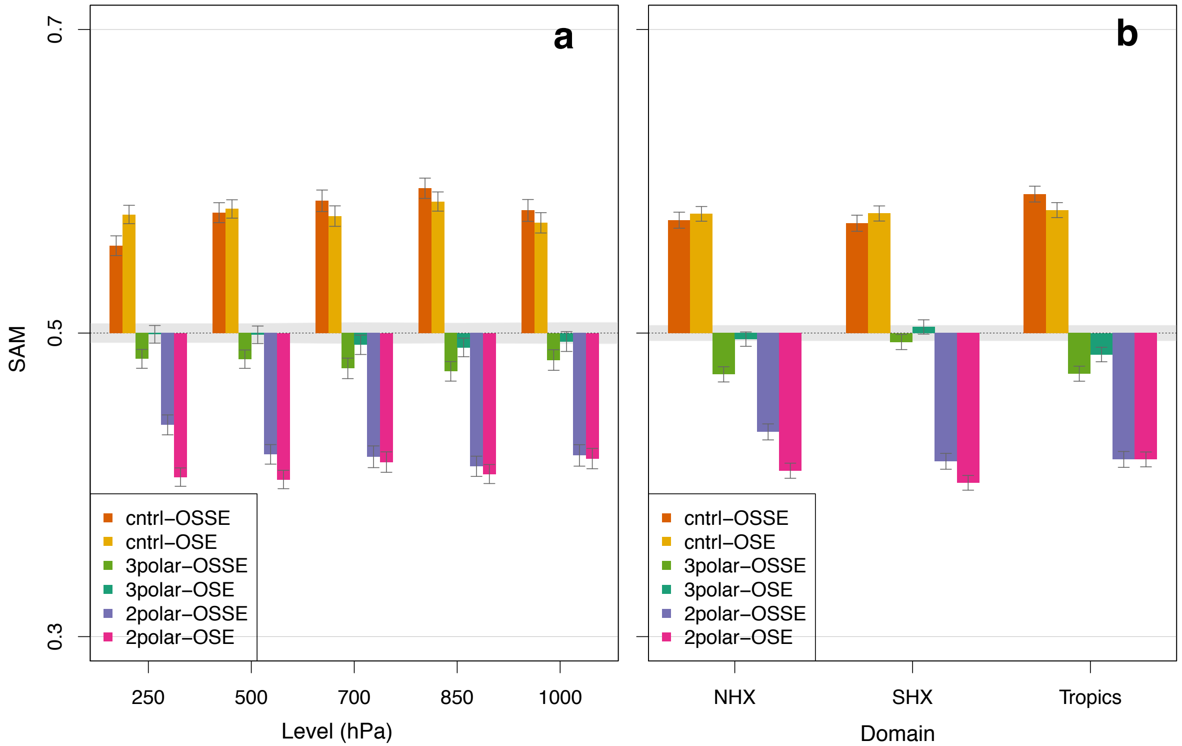


- All data gap scenarios result in poor forecast skills
- Tendency of impact mostly as expected although there are bit of variabilities in OSSE vs OSE inter-comparison results

Results: SAMs global and vs. forecast time



Results: SAMs by level and domain



more…

- email:
  - ross.n.hoffman@noaa.gov
- Oct 2018 JTECH paper:
  - doi: 10.1175/JTECH-D-18-0061.1
- AMS presentation:
  - https://ams.confex.com/ams/2019Annual/meetingapp.cgi/Paper/353842